Attachment 1d: Copy of Document 1 from the Springer book

Bolt, R. (1980) Put that there: Voice and Gesture at the Graphic Interface, *Computer Graphics, 14(3)*, pp. 262-270.

Cohen, M., Murveit, H., Bernstein, J., Price, P., and Weintraub, M. (1990) The DECIPHER Speech Recognition System. In *1990 IEEE ICASSP*, pp. 77-80.

Cohen, P. (1992) The role of natural language in a multimodal interface. In *Proceedings of UIST'92*, pp. 143-149.

Cohen, P.R., Cheyer, A., Wang, M. and Baeg, S.C. (1994) An Open Agent Architecture. In *Proceedings AAAI'94 - SA*, Stanford, pp. 1-8.

*Dauphin DTR-1 User's Manual*, Dauphin Technology, Inc., Lombard, Ill 60148.

Faure, C. and Julia, L. (1994) An Agent-Based Architecture for a Multimodal Interface. In *Proceedings AAAI'94 - IM4S*, Stanford, pp. 82-86.

Genesereth, M. and Singh, N.P. (1994) *A knowledge sharing approach to software interoperation*, unpublished manuscript, Computer Science Department, Stanford University.

*Telescript Product Documentation* (1995), General Magic Inc.

Koons, D.B., Sparrell, C.J., and Thorisson, K.R. (1993) Integrating Simultaneous Input from Speech, Gaze and Hand Gestures. In *Intelligent Multimedia Interfaces*, Maybury, M.T. (ed.), Menlo Park: AAAI Press/MIT Press.

Maybury, M.T. (ed.) (1993) *Intelligent Multimedia Interfaces*, Menlo Park: AAAI Press/MIT Press.

Neal, J.G., and Shapiro, S.C. (1991) Intelligent Multi-media Interface Technology. In *Intelligent User Interfaces*, Sullivan, J.W. and Tyler, S.W. (eds.), Reading: Addison-Wesley Pub. Co., pp. 11-43.

Nigay, L. and Coutaz, J. (1993) A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In *Proceedings InterCHI'93*, Amsterdam, ACM Press, pp. 172-178.

Object Management Group (1991) *The Common Object Request Broker: Architecture and Specification*, OMG Document Number 91.12.1.

Oviatt, S. (1994) Toward Empirically-Based Design of Multimodal Dialogue Systems. In *Proceedings of AAAI'94 - IM4S*, Stanford, pp. 30-36.

Oviatt, S. and Olsen, E. (1994) Integration Themes in Multimodal Human-Computer Interaction. In *Proceedings of ICSLP'94*, Yokohama, pp. 551-554.

Park, S.K., Choi J.M., Myeong-Wuk J., Lee G.L., and Lim Y.H. (submitted for publication), *MASCOS : A Multi-Agent System as the Computer Secretary*.

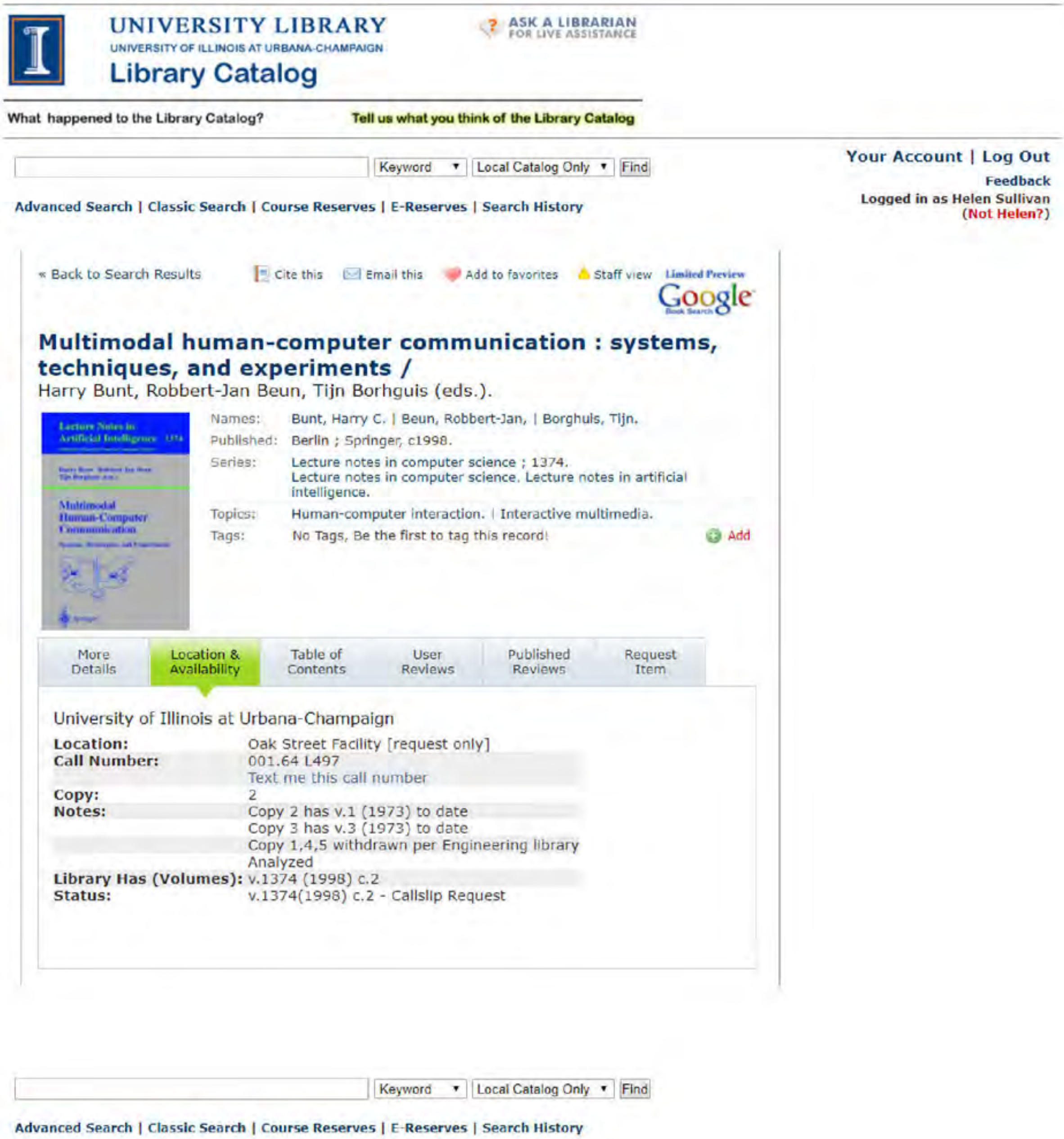
Rhyne J. (1987) Dialogue Management for Gestural Interfaces, *Computer Graphics, 21(2)*, pp. 137-142.

Schwartz, D.G. (1993) *Cooperating heterogeneous systems: A blackboard-based meta approach*, Technical Report 93-112, Center for Automation and Intelligent Systems Research, Case Western Reserve University, Cleveland Ohio, (unpublished PhD. thesis).

Sullivan, J. and Tyler, S. (eds.) (1991) *Intelligent User Interfaces*, Reading: Addison-Wesley Pub. Co.

Warren, D. and Pereira, F. (1982) An Efficient Easily Adaptable System for Interpreting Natural Language Queries, *American Journal of Computational Linguistics, 8(3)*, pp. 110-123.

# Attachment 1e: University of Illinois at Urbana-Champaign Library catalog record for the Springer book



**Multimodal human-computer communication : systems, techniques, and experiments /**

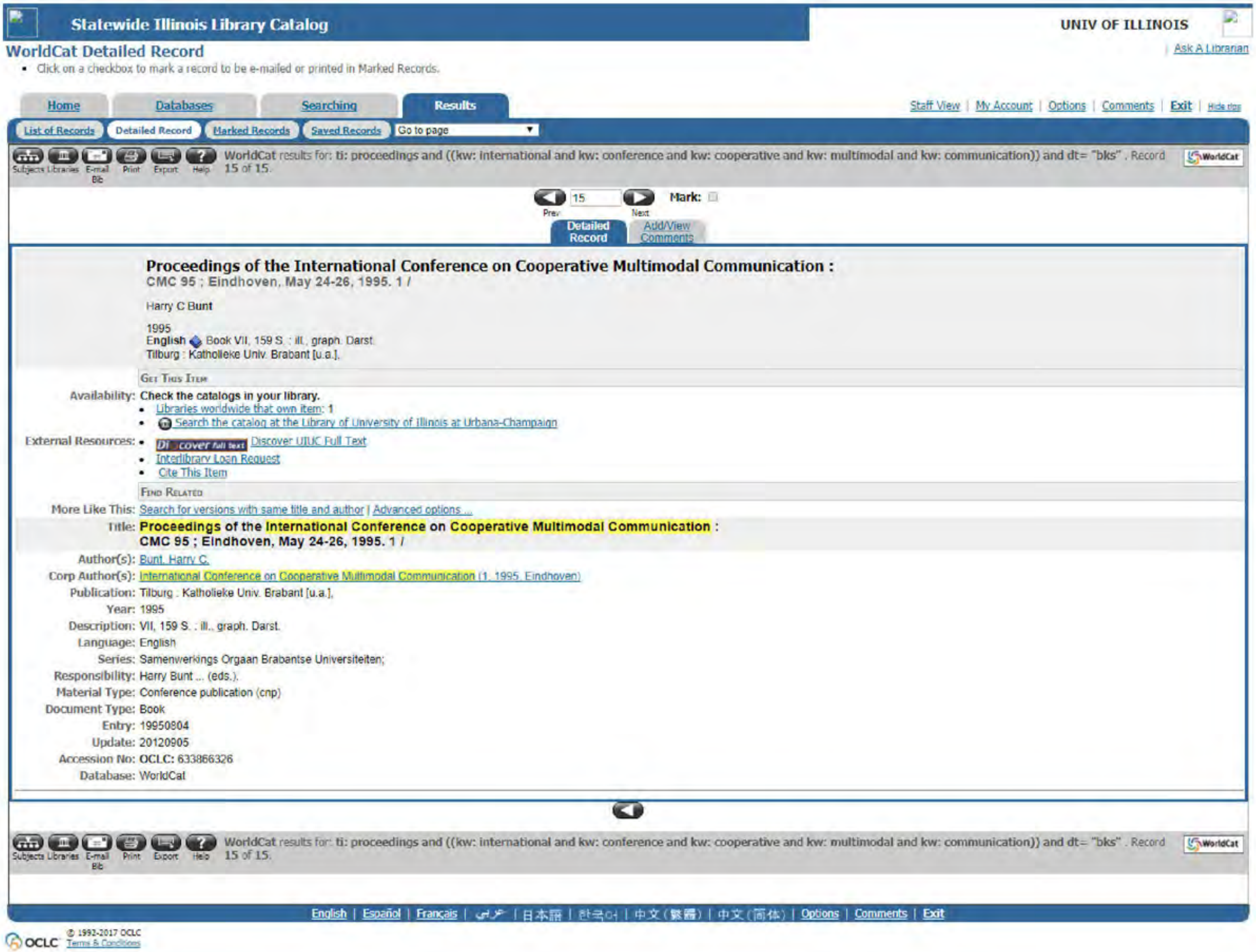
Harry Bunt, Robbert-Jan Beun, Tijn Borhguis (eds.).

| | |
|---|---|
| Names: | Bunt, Harry C. \| Beun, Robbert-Jan, \| Borghuis, Tijn. |
| Published: | Berlin ; Springer, c1998. |
| Series: | Lecture notes in computer science ; 1374.<br>Lecture notes in computer science. Lecture notes in artificial intelligence. |
| Topics: | Human-computer interaction. \| Interactive multimedia. |
| Tags: | No Tags, Be the first to tag this record! |

University of Illinois at Urbana-Champaign

| | |
|---|---|
| **Location:** | Oak Street Facility [request only] |
| **Call Number:** | 001.64 L497<br>Text me this call number |
| **Copy:** | 2 |
| **Notes:** | Copy 2 has v.1 (1973) to date<br>Copy 3 has v.3 (1973) to date<br>Copy 1,4,5 withdrawn per Engineering library<br>Analyzed |
| **Library Has (Volumes):** | v.1374 (1998) c.2 |
| **Status:** | v.1374(1998) c.2 - Callslip Request |

# Attachment 1f: Statewide Illinois Library Catalog record for the CMC/95 proceedings



**WorldCat Detailed Record**

- Click on a checkbox to mark a record to be e-mailed or printed in Marked Records.

WorldCat results for: ti: proceedings and ((kw: international and kw: conference and kw: cooperative and kw: multimodal and kw: communication)) and dt= "bks" . Record 15 of 15.

15

**Proceedings of the International Conference on Cooperative Multimodal Communication :**
**CMC 95 ; Eindhoven, May 24-26, 1995. 1 /**

Harry C Bunt

1995
English Book VII, 159 S. : ill., graph. Darst.
Tilburg : Katholieke Univ. Brabant [u.a.],

Availability: **Check the catalogs in your library.**
- Libraries worldwide that own item: 1
- Search the catalog at the Library of University of Illinois at Urbana-Champaign

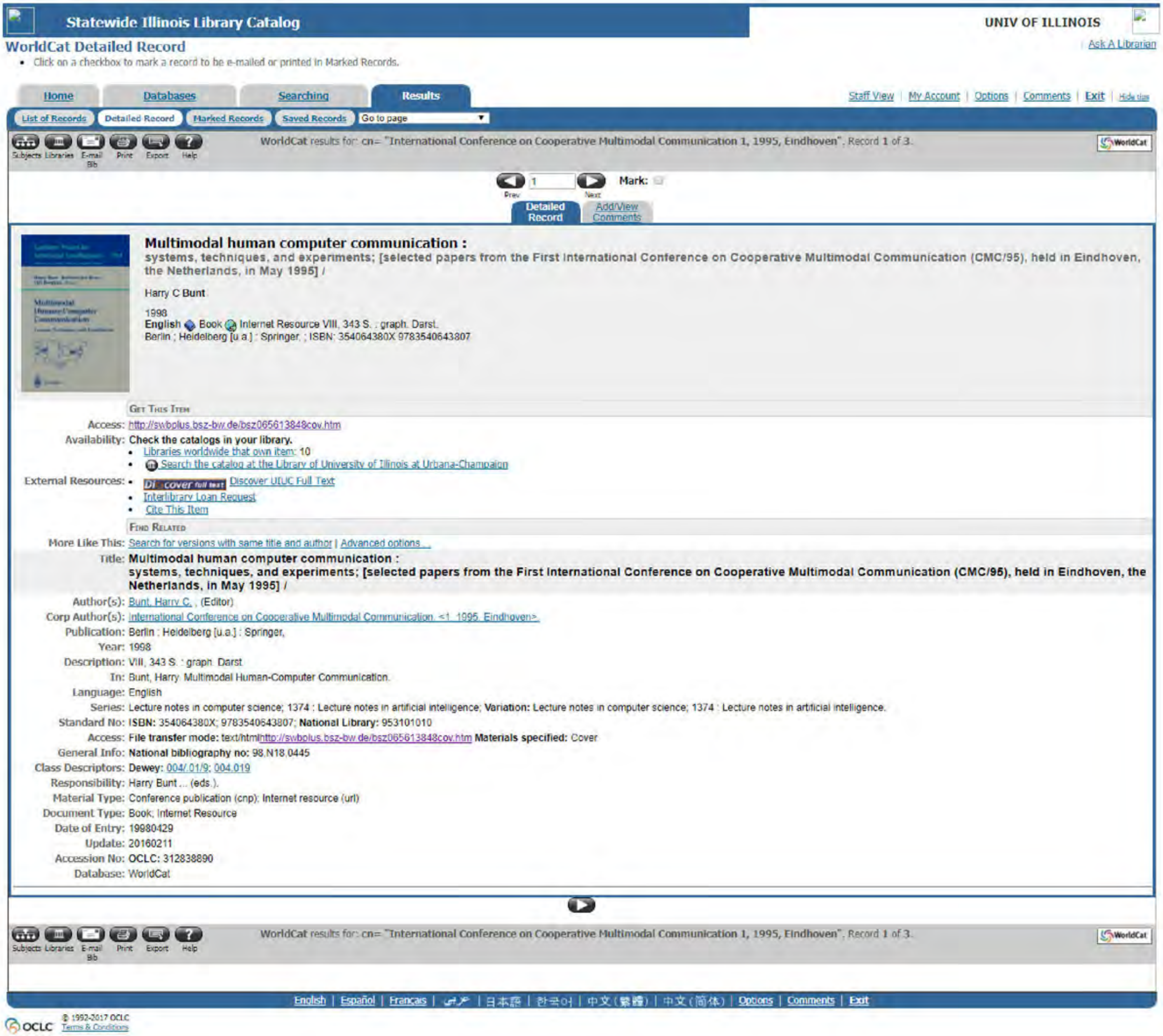
External Resources:
- Discover UIUC Full Text
- Interlibrary Loan Request
- Cite This Item

More Like This: Search for versions with same title and author | Advanced options ...

Title: **Proceedings of the International Conference on Cooperative Multimodal Communication : CMC 95 ; Eindhoven, May 24-26, 1995. 1 /**
Author(s): Bunt, Harry C.
Corp Author(s): International Conference on Cooperative Multimodal Communication (1, 1995, Eindhoven)
Publication: Tilburg : Katholieke Univ. Brabant [u.a.],
Year: 1995
Description: VII, 159 S. : ill., graph. Darst.
Language: English
Series: Samenwerkings Orgaan Brabantse Universiteiten;
Responsibility: Harry Bunt ... (eds.).
Material Type: Conference publication (cnp)
Document Type: Book
Entry: 19950804
Update: 20120905
Accession No: OCLC: 633866326
Database: WorldCat

WorldCat results for: ti: proceedings and ((kw: international and kw: conference and kw: cooperative and kw: multimodal and kw: communication)) and dt= "bks" . Record 15 of 15.

# Attachment 1g: Statewide Illinois Library Catalog record for the Springer book



**WorldCat Detailed Record**

- Click on a checkbox to mark a record to be e-mailed or printed in Marked Records.

WorldCat results for: cn= "International Conference on Cooperative Multimodal Communication 1, 1995, Eindhoven". Record 1 of 3.

**Multimodal human computer communication :**
**systems, techniques, and experiments; [selected papers from the First International Conference on Cooperative Multimodal Communication (CMC/95), held in Eindhoven, the Netherlands, in May 1995] /**

Harry C **Bunt**

1998
**English** Book Internet Resource VIII, 343 S. : graph. Darst.
Berlin ; Heidelberg [u.a.] : Springer. ; ISBN: 354064380X 9783540643807

GET THIS ITEM

Access: http://swbplus.bsz-bw.de/bsz065613848cov.htm
Availability: **Check the catalogs in your library.**
- Libraries worldwide that own item: 10
- Search the catalog at the Library of University of Illinois at Urbana-Champaign

External Resources:
- Discover full text Discover UIUC Full Text
- Interlibrary Loan Request
- Cite This Item

FIND RELATED

More Like This: Search for versions with same title and author | Advanced options ...
Title: **Multimodal human computer communication : systems, techniques, and experiments; [selected papers from the First International Conference on Cooperative Multimodal Communication (CMC/95), held in Eindhoven, the Netherlands, in May 1995] /**
Author(s): Bunt, Harry C. , (Editor)
Corp Author(s): International Conference on Cooperative Multimodal Communication <1 1995 Eindhoven>.
Publication: Berlin : Heidelberg [u.a.] : Springer,
Year: 1998
Description: VIII, 343 S. : graph. Darst.
In: Bunt, Harry. Multimodal Human-Computer Communication.
Language: English
Series: Lecture notes in computer science; 1374 : Lecture notes in artificial intelligence; **Variation:** Lecture notes in computer science; 1374 : Lecture notes in artificial intelligence.
Standard No: **ISBN:** 354064380X; 9783540643807; **National Library:** 953101010
Access: **File transfer mode:** text/htmlhttp://swbplus.bsz-bw.de/bsz065613848cov.htm **Materials specified:** Cover
General Info: **National bibliography no:** 98,N18,0445
Class Descriptors: **Dewey:** 004/.01/9; 004.019
Responsibility: Harry Bunt ... (eds.).
Material Type: Conference publication (cnp); Internet resource (url)
Document Type: Book; Internet Resource
Date of Entry: 19980429
Update: 20160211
Accession No: OCLC: 312838890
Database: WorldCat

WorldCat results for: cn= "International Conference on Cooperative Multimodal Communication 1, 1995, Eindhoven". Record 1 of 3.

or to spatial disposition in the sector), the radar, the telephone, the displays. The strips, the radar display, and listening to radio and telephone communications, appear to represent important communicative resources.

## 5 Methodology

We have analysed how air traffic controllers solve conflicts. A simulated situation of air traffic control was composed of six conflicts evolving over a period of about 30 minutes. We videotaped four teams of two controllers who had to regulate this traffic (the executive controller in charge of the traffic regulation of communication with pilots and the planning controller).

Verbal and non verbal communications were analysed in terms of exchanges. We consider as exchanges one or many interventions focusing on the same object. Each intervention corresponding to turn talking or non-verbal action, is considered here as a basic unit of communication. In this analysis, we distinguished four types of interventions corresponding to (1) verbal communication between controllers (2) verbal communication between executive controller and pilots (3) non-verbal communication related to strip handling (writing, moving or pointing to strips) and (4) non-verbal communication related to radar scope (pointing to radar scope). Three principles guide the collection of observations concerning the non-verbal element of the interaction, these are: (1) the nature of the observations selected, (2) their description in reference to the activity, and (3) their processing in relation to the action under way. Taking into account the extreme richness of non-verbal or para-verbal elements, we focused on the behaviour related to the use of available environmental resources in the control environment, and on the content of verbal communications (for example, the handling of strips, pointing to the radar scope). In addition, all these elements are described with a set of action verbs (e.g. "write", "give", "place", "point", "take", "shift", "move", etc.) The characteristics of this list are the observable nature of the selected behaviour (actions which can be described using the professional vocabulary) and the coherence of the list vis-a-vis the control activity (e.g. "shift" the strip out of line on the board has a distinctly different meaning to "move" the strip). Amongst the non-verbal actions taken into account are (a) deictics (indicating an object or some specific information such as flight level on the strip, possibly asssociated with comments such as "this one", "here", "th! ere", and other illustrative gestu res, (b) writing (various annotations on the strip are made by controllers), and (c) handling of objects (the strip board is organized by moving the strips on the columns or between them).

## 6 Results

We now intend to develop the points developed above through the analysis of several examples. The following abbreviations will be used: (Cp (planning controller), Ce (executive controller), Pi (pilot).

### 6.1 Non-verbal resources as a specific means of communication

Several examples illustrate the specific role of non-verbal resources in communicating: non-verbal resources are revealed to be a specific vector of communication for some types of information which are not verbally expressed (e.g. urgency of the situation).

The planning controller receives the strips (several minutes before the aircraft effectively enters the sector) which he transmits to the executive controller The gestures associated with the transmission of the strip is "multiform" (passing the strip from hand to hand, placing it on the edge of the board, using the strip to point to the radar scope, etc.). The strip can also be used for annotations (marking the evolution of the flight, or circling the destination airport).

Transmission of strip according to the non-routine character of the information

Team 4 - V
(11)
*Cp-Pi:"IBERIA takeoff, IBERIA 913 to Amboise OK?" (Cp writes strip, places strip on edge of board)*
Ce-Cp:"OK" (Ce writes strip, keeps strip in his hand)
Ce -Cp:"How many does he want that man there, he wants 330 too."
(...)
Ce-Pi: "IBERIA 913 initial level 280 " (Ce writes strip, holds Iberia913 strip)
(...)
(Cp presents Other strip)
(12)
(...)
(Ce holds Iberia913 strip)
(Cp holds Other strip)
(...)
(Ce puts IBE.913 strip down, picks up Other strip, places on right board)
(...)
(Ce picks up Iberia913 strip again) Ce-Cp "IBERIA..."
Cp-Ce: "hey?"
Ce-Cp: "913 I'll put him straight onto Amboise, why not? He is with us"
Cp-Ce: "From now on you'll be all right"
Ce-Cp: "He is with us" (Ce points to Scope)
Ce-Pi: "IBERIA 619 maintain level"
(Ce points to Iberia913 strip. Ce underlines Iberia613 strip.)
Ce-Pi: "IBE913 turn left to AMB" (Ce holds Iberia913 strip)
(13)
Ce-Cp: "and in any case he can only have 290 huh? En 330 it's not possible?"
Cp-Ce: "who's that, the 913? why not?"
Ce-Cp: "330 on amboise, 290 is not possible"
(Ce points to scopeX2) (Cp places Iberia913 strip)
Cp-Ce: "there's nothing there... I'll sort you out, Airbus 300.. and the other at 330, all right we'll sort it out he said... Bordeaux is taking everything for once"
(Cp points to strips, Iberia913 strip and Monarque 598 strip)
(...)

In this example, it would appear that the planning controller transmits Iberia913 strip by placing it sideways on the edge of the board in front of the executive controller. The planning controller associates a verbal intervention with two non-verbal ones (underlining the strip and placing it on the board). The results of these actions are first to improve memorization of the requested level, second to highlight for his/her colleague, a conflict with another aircraft in the same sector at the same level. The act of it placing the strip sideway on the board can be interpreted by the executive controller as a mark of non-routine information which has to be taken into account quickly (the strip is not placed outside the board which forces him to pick it up to visualise the strips which are integrated in the board).

In this same example, it appears that gestures associated with strip handling can be interpreted by co-workers; here the executive controller's activity is constrained by communication with the pilots (not reported here), he keeps the strip in his hand several minutes (approximately three minutes) before looking to solve this problem which is the purpose of an exchange with the planning controller. A second strip is proposed by the planning controller (Other strip), he will have to delay transmission to the executive controller who kept the Iberia 913 strip without putting it on the board.

In the following example, non-verbal resources appears to manifest pieces of information which are not verbally expressed. First, the executive controller points to the scope it with a turning movement which represents a regulation strategy.

Conflict solving expressed by gesture

Team 2-VI
(20)
Cp-Ce: "Yes, he's coming, he's here" (Cp points to ACF5111 strip)
Ce-Pi: "AIR CHARTER 5111, good morning, 26 at Orly maintain 270 call back for descent."
(Ce writes ACF5111 strip)
Ce-Cp: "430 knots, 430 knots, 430 knots, it's going to be easy"
(Ce points to X3 scope)
Cp-Ce: "You know it's an Airbus"
(...)
Cp-Ce: "It's..."
Ce-Cp: "One fare... as quiet as that, great"
*(Ce points to scope, turning movement)*
Cp-Ce: "yes"
Ce-Cp: "Go on, climb"
(...)

## 6.2 Complementary role in communication regulation

Non-verbal resources are used in giving a context of interpretation for what is verbally expressed. In this analysis, our second interest is to examine the role they play in the success of communication. The study of multimodal aspects of communication leads us to underline the importance of non-verbal acts to understand communications, as part of the context activated to interpret communications. In the following example, the exchange is based on the establishment of the referenced flight, two aircraft are present in the sector, SWR011 and SWR012. The executive controller's hypothesis is that SWR concerned is SWR012, the planning controller's interventions, both indicating the strip and commenting verbally, allow him to make his intention explicit.

Strip pointing in communication regulation

Team 3 -III
(13)
Cp-Ce: "watch out for the SWR eh?"
(...)
(14)
(...)
Cp-Ce: "We can transfer *the SWR* on frequency already"
(Ce wants to take SWR012 strip) Ce-Cp : the SWR012"
(Cp points to SWR012 strip) *Cp-Ce: "no, not that one"*
Ce-Cp: "all right,*the SWR011* you mean at 280"
Cp-Ce:"no, no *...011" (Cp points to SWR011 strip)*

This example underlines the richness of the multimodal communication which allows determination of the object which is referred to and to ensure success of communication (the refrence SWR" is first based on a misunderstanding between controllers; in the second part, reference to SWR011 is expressed four times). In the following example, two fights are referred to, Monarque 598 and Monarque 1789, both are the subject of a conflict solving. What is interesting here is that the controllers never say explicitly "598" or "1789", but pointing to the radar scope allows them to confirm that the planning controller's first hypothesis is not relevant.

Pointing at scope in communication regulation

Team 4-III
(10)
(...)
Ce-Cp: "I turned *the Monarque a bit too far"*
Cp-Ce: "the Monarque? no, it's fine"
Ce-Cp: "no, him"
(Ce points to the scope)
Cp-Ce:"ah, that one"
Ce-Cp: "the MON is just right.
(...)

As shown in the above two examples, several misunderstandings arose in communication between controllers. Verbal and non-verbal resources appear complementary to ensure the establishment of a mutual cognitive environment. This complementarity is not only used in the case of misunderstandings as we will see in the following examples.

In the next example, the executive controller clarifies the reference by pointing to aircraft AFR022 on the radar scope. The planning controller confirms his identification of the aircraft in both a verbal and non-verbal way.

Verbal interventions combined with strip moving and pointing at radar scope in diagnosis.

Team 2-1
(04)
*Ce-Cp: "Right, there's one there." (Ce points at scope)*
*Cp-Ce: "Yes, it's AFR 022 I think" (Cp points to Afr022 strip)*
(...)
Ce picks up Afr022 strip and puts it in 5 shifted out of line on the left)
Ce : "things are getting heated there, it's not reasonable!"

In this example, reference is made twice, first by moving the strip on the board and secondly by pointing at the radar scope. The request concerns remembering an aircraft to be transferred to an adjacent sector. The name of the aircraft is not verbally expressed .

Verbal interventions combined with moving the strip and pointing at the radar scope.

Team 3-11
(...)
(20)
(...)
Ce-Cp: "that one there, I'll transfer his frequency a bit later, will you remind me? he's there"
(Cp removes Monarque598 strip, places it on right hand board, points to scope)
Cp-Ce: "all right"
(...)

In a similar way, overhearing the radio frequency allows the planning controller to initiate a request based on the reference made to the communication between the executive controller and the pilot. This next example shows the way in which controllers use environmental resources (here overhearing a radio communication) to initiate an exchange. What is interesting here is that this is an economical way to communicate (reference made to radio communication). This would probably have consequences on mutual understanding in the sense that the cognitive effort required to interpret the meaning of the utterance*(Cp -Ce: "which way are you sending him?")* stays less than if the controller had referred to another aircraft in the sector, indeed, in respect to the relevance principle (Sperber, Wilson, 1989), the utterance expressed by the planning controller is relevant in the cognitive environment of the interlocutor.

To conclude on this point, it would appear that communications between controllers are usually composed of both verbal and non-verbal resources in one single exchange. The results (obtained through videotaping of four teams) reveal that the exchanges where non-verbal resources are used, are the most numerous. Exchanges composed of only verbal interventions are (percentage): 24.61 (team 1), 25.73 (team 2), 21.55 (team 3) and 28.63 (team 4). Exchanges composed of interventions which are either verbal and non-verbal, or only non-verbal, are 75.37 (team 1), 74.51 (team 2), 78.42 (team 3), and 71.34 (team 4). This second category of exchanges includes three different types of exchange: (a) use of only non-verbal communication (between 6 and 16 depending on the team, considering the total number of exchanges); (b) use of non-verbal communication supported by strip handling combined with use of verbal communication that is here radio-communication with the pilot (between 15 and 35); (c) use of both verbal and non-verbal communication (between 28 and 51).

Overhearing radio frequency with verbal intervention

Team 4 - III
(09)
*Ce-Pi: "LUFT 498 say heading..."*
*Cp-Ce: "which way are you sending him?"*
Ce-Cp: "to the right"
Ce-Pi: "LUFT498 turn right heading 270"
Ce-Cp:"where did I put it?"
(...)
Cp-Ce:"haven't you got it?"
(...)
Cp-Ce: "You've got the LUFT which is there"
Cp points to DLH498 strip
Ce-Cp: "270"
(Ce writes DLH strip, holds it)
(etc...)

## 6.3 Verbal and non-verbal resources in dialogue management

This analysis of non-verbal acts leads us to consider, on the one hand, the action with communicative intention (for example, handling the strip and simultaneously asking something to notify relevant information), and on the other hand the action considered as a means of organizing information for him/herself (these actions can be interpreted by co-workers as intention recognition). In other words, informative actions and utterances (versus communicative) serve as a basis for inferences that agents make about their mutual cognitive environment, but are not necessarily intended to communicate. The co-presence in the same working location allows the construction of mutual beliefs in ways which are efficient, in the sense that they are not intrusive in comparison with the verbal channel.

In the following example, the controllers initiate a common diagnosis of the situation, the planning controller acts on the strip in writing and moving the strip on the column on the flight progress board. Two minutes later, the executive controller's first intervention can be interpreted by the planning controller as confirmation of his diagnosis concerning a problem at level 330 including DAN4446. One minute before the planning controller has given the strip for AFR022 which is descending in the opposite direction. In analysing the situation, the executive controller shows that he does not find the flight on the radar scope, by tapping with his finger on the strip he is referring to, which is followed by the planning controller's engagement in this exchange. Therefore, such a non-verbal act can be seen as an implicit request in the sense that the planning controller is supposed to answer only if he is not engaged in an activity which is difficult to interrupt.

Strip pointing while forming a common diagnosis of the situation

Team 2-1
(00)
Cp-Ce: "There's the 330" (Cp underlines strip Dan1116)
Ce-Cp: "Well yes... it's going to be hard"
(Cp shifts Dan1116 to the right)
Cp-Ce: "We'll have to see"(...)
(01)
(Cp gives two strips, Afr022 strip and FBJMG strip)
Cp -Ce: "Here's a problem."
Ce: "There are lots of problems." (Ce takes slips, places one in pos.4 AFR022, shifts it to the left and keeps the other in his hand)
Ce "Ah yes"
(02) (...)
Ce-Cp: "Yes there seems to be a lot of climbing to 330"
*(Ce points to Dan4446 strip, tapping it with his finger)*
Ce-Cp: "That's the one I can't see."
Cp-Ce: "Just a moment, yes, he is behind."
(Cp adjusts scope, points to scope)
(...)

In the following example, it appears that the executive controller does not explicitly ask the planning controller for assistance and does not verbally express the name of the aircraft concerned. The planning controller listens to the pilot and is able to point to the flight progress strip. After the pilot's call, the executive controller looks for the DAN4446 strip, this strip had been put on the board a few minutes before, further to the analysis made previously. This position is no longer in accordance with the current situation. The planning controller, hearing the radio-communication, can infer who "this one" is, and points to the associated strip. Thus, the reference is made through radio-communication and confirmed by the planning controller twice by pointing and verbal intervention. We note that the executive controller never named the aircraft explicitly. Radio-communications between the executive controller and pilots, preceded or followed by strip handling can be considered as an important part of the controller's work. Strip handling allows the controller to organize and process information related to each flight according to his/her diagnosis or actions on conflicts.

Overhearing radio frequency with verbal interventions and moving strips

Team 2-1
(06)
*Pi-Ce: "..."*
*Ce: "I can't find him any more, that one"*
Ce-Pi: "Who was calling?" Pi-Ce: "DAN4446 Good morning"
*(Cp points to Dan4446 strip)*
Ce -Pi: "please maintain level 290 I'll call you back ... it's not right time not" (Ce underlines Dan 4446 strip)
Ce-Cp: "Wrong place" (Ce places Dan 4446 strip shifted out of line to the right)
Cp-Ce: "Yes, you'd put it there because of the conflict..."
(07)
Ce-Cp : "Shhh! ... Just a minute ..."
Ce-Pi: "AFR022 Good morning it's facing west at Roissy - please descend to flight level 200 ... two zero zero." (Ce writes Afr022 strip)(...)

## 7 Discussion

The study of cooperation between human agents in such face-to-face situations, emphasizes the role of external artifacts in cooperation processes because they allow the cooperative agents to organize their own cognitive processes (external artifacts are used as support for memory and problem solving) at the same time as updating their mutual cognitive environment through intention recognition processes. Intention recognition through non-verbal communication can probably be seen as essential for cooperation for several reasons. First, non-verbal communication allows agents to communicate elements like urgency, etc. which are not explicity verbalized. Second, it appears that a large amount of verbal communications are associated to non-verbal ones; in this sense they constitute a context of interpretation of what is verbally expressed, which is used to regulate misunderstandings. Third, its non-interruptive property appears important(instead of making explicit verbal requests! , each agent notices non-verbal acts by other agents as part of their cognitive activities). In the perspective of the design of cooperative tools, difficulties arise from developing methodologies to anticipate and evaluate the implications of a new environment on the cooperation processes between agents. The design of complex working environments is conducted today by using the following methodologies: (1) validation, based on empirical approaches, use of prototypes and iterative testing (Gaillard and Leroux, 1994), (2) simulation, but methodological tools for the anticipation of cognitive activities from actual situations have to be investigated. The first step of this is to assess the cooperative nature of the working environment. This is based on the analysis of how people use external artifacts not only in order to support their own cognition, but also to cooperate on implicit modes. The methodology is based on assessment of the capacity of external artifacts to support intention recognition and includes two stages: (1) identifying in a working situation, the ! supports used for intention recognition and for cognition (this analysis highlights how all external artifacts are used in diagnosis, problem solving, etc.), in nominal and degraded situations (Bressolle, 1992); (2) anticipate or evaluate how these cognitive supports are transferred during the introduction of new technologies in relation to the cognitive activities in the domain. In this second stage, in order to anticipate cognitive properties of new working environments, a simulation tool is being developed based on the formalisation and simulation of the communications between cooperative agents and on conceptual specifications of the application (Zorola-Villarreal et al. 1995).

## References

Benchekroun H., Pavard B., Salembier P. [1993] Design of cooperative systems in complex dynamic environments. In J.M. Hoc, P.C. Cacciabue and E. Hollnagel (eds.) Expertise and Technology - Cognition and Human-Computer Cooperation. 16p.

Bressolle M. C. [1992] Perception de l'intention et cooperation dans le cas du controle de la navigation aerienne. Premiers elements d'analyse. Internal report.

Cadoz [1993] Le geste canal de communication homme-machine, la communication instrumentale. Ecole d'Ete ARC/PRC CHM. Communication et Multimodalite dans les Sytemes Naturels et Artificiels. Chateau de Bonas, France. 35- 67.

Clark H. H. and Wilkes-Gibbs D. [1990] Referring as a collaborative process. In Intention in Communication. Cohen P.R., Morgan J., Pollack M.E.(eds). Massachussets Institute of Technology. System development Foundation Benchmark series. The MIT press. Cambridge. 463- 493.

Clark H. H. and Schaeffer E. F. [1989] Contributing to discourse. Cognitive Science. 13. 259-294.

Clark H. H. and Brennan S. E. [1991] Grounding in conversation. In Perspective on Socially Shared Cognition. Resnick, Levine and Teasley (eds). Washington DC.127-149

Cosnier J., Brossard A. [1984] Communication non-verbale : co-texte ou contexte? In La Communication Non-verbale. J. Cosnier and A. Brossard (eds). Delachaux et Niestle. 1-29.

Dourish P. and Belloti V. [1992] Awareness and coordination in shared workplaces. In Proceedings of ACM Conference on Computer Supported Cooperative Work. Toronto, Canada. 107-114

Gaillard I., Leroux M. [1994] Improving air traffic control : proving new tools or approving the joint human- machine system? In Human Factors Certification of Advanced Aviation Technologies. J.A. Wise, V.D. Hopkin and D.J. Garland (eds). Aviation Human Factors series. 275-287.

Heath C. and Luff P. [1991] Collaborative activity and technological design : task coordination in London underground control rooms. In Proceedings of the Second European Conference on Computer-Supported Cooperative Work (ECSCW'91). L. Bannon, M. Robinson and K. Schmidt (eds). 65-80.

Heath C., Jirotka M., Luff P. and Hindmarsch J.[1993]Unpacking collaboration : the Interactional Organisation of Trading in a Dealing Room. Proceedings of the Third European Conference on Computer Supported Cooperative Work.G. de Michelis, C. Simone and K.. Schmidt (eds) 13-17 Sept, Milan, Italy.155-170

Hughes J.A., Randall D. and Shapiro D. [1992] Faltering from Ethnography to design. In ACM 1992 Proceedings of the Conference on Computer-Supported Cooperative Work. 115-122.

Hutchins E. and Klausen T. [1992] Distributed cognition in an Airline Cockpit. Report. Departement of Cognitive Science. University of California, Institute of Psychology.

Krauss R. M. and Fussel S. R. [1990] Mutual knowledge and communicative effectiveness. In Intellectual Teamwork, Social and Technological Foundations of Cooperative Work. J. Galegher, R. Kraut, C. Egido (eds). 111-145.

Leroux M. [1992] The role of verification and validation in the design process of knowledge based components of air traffic control systems. In Verification and validation of Complex and Integrated Human - Machine Systems. J.A. Wise, V.D. Hopkin and P. Stager (eds). Vimeiro, Portugal : Nato Advanced Study Institute.

Rogers Y. [1992] Ghosts in the network : distributed troubleshooting in a shared working environment. Proceedings Conference on Computer Supported Cooperative Work. November 1992, Toronto, Canada. 346-355.

Shapiro D.Z., Hughes J.A., Randall D., Harper R. [1989] Visual Re-representation of database information - The flight data strip in air traffic control. Report. Lancaster University. Department of Sociology, Lancaster University, UK.

Scherer K. R. [1984] Les fonctions des signes non-verbaux dans la conversation. In La communication Non-verbale. J. Cosnier and A. Brossard (eds). Delachaux et Niestle. 71-100.

Sperber D. and Wilson D. [1989] La pertinence - Communication et cognition. Ed de Minuit. (trad Relevance Communication and Cognition, 1986). 393p.

Zorola-Villarreal R., Pavard B., Bastide R. [1995] SIM-COOP : a tool to analyse and predict cooperation in complex environments - A case study : the introduction of a datalink between controllers and pilots. Submitted to Fifth International Conference on Human-Machine Interaction and Artificial Intelligence in Aerospace. Toulouse, France, Sept 1995.

# Cooperative Multimodal Communication in the DenK Project

Harry Bunt[1], René Ahn[1], Robbert-Jan Beun[2], Tijn Borghuis[3] & Kees van Overveld[3]

[1] Institute for Language Technology and Artificial Intelligence (ITK)
Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.
[2] Institute for Perception Research (IPO)
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
[3] Faculty of Mathematics and Computing Science Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
e-mail: Harry.Bunt@kub.nl

### Abstract

In this paper we present the DenK-project, which aims at building a generic cooperative human-computer interface combining linguistic and visual interaction. We discuss the basic principles underlying the project and the emerging DenK-system. The system integrates results from fundamental research in knowledge representation, communication, natural language semantics and pragmatics, and object-oriented animation. Our design incorporates a cooperative and knowledgeable electronic assistant that communicates with a user in natural language, and an application domain which is presented visually to the user. The assistant, that we call the *cooperator*, has an information state that is represented in a rich form of type theory, a formalism that enables us to model the inherent cognitive dynamics of a dialogue participant. Pragmatic issues in man-machine interaction, concerning the use of natural language and knowledge in cooperative communication, are central to our approach.

**Key words:** multimodal interaction, knowledge representation, natural language semantics, pragmatics, type theory, context modelling, object-oriented animation

## 1 Introduction

The DenK project is long-term collaborative research activity of the universities of Tilburg and Eindhoven, started in 1989 and expected to run until 1998.[1] The project aims at the exploration, formalization and application of fundamental principles of communication from a computational perspective, in order to build advanced cooperative human-computer interfaces. It gives a central position to the formal modelling of dialogue mechanisms and dialogue contexts, with particular emphasis on the role of shared knowledge.

The program combines fundamental research in knowledge representation, communication, natural language semantics pragmatics, and object-oriented animation. Techniques from these domains are applied in the prototypical *DenK-system*, the design of which reflects

---

[1] 'DenK' is an abbreviation of 'Dialoogvoering en Kennisopbouw', which means 'Dialogue Management and Knowledge Acquisition'. The word *denk* in Dutch means *think*.

a situation where two participants cooperate who can exchange information about an application domain they can both observe, and in which they can both manipulate objects through direct manipulation. The system can 'observe' the application domain in the sense that it incorporates a formal model of the current state of the domain as well as a description of its visualizable aspects, which it can consult. The user can observe the domain by looking at its visualization on the screen, and can use the mouse to directly manipulate objects in this visualization. The system can 'directly manipulate' the domain in the sense that it can internally generate and execute commands to create, change or delete objects.

Point of departure in the DenK-project is that, from a user's point of view, a computer should ideally present itself as an intelligent 'electronic assistant' who is knowledgeable about the task and the domain of the application. The electronic assistant should interact in an intelligent and cooperative way with the user, using linguistic and visual modalities as appropriate, and acting on the user's intentions as understood in the context of the interaction. The DenK-system is intended to be *generic* in that its architecture, as well as many of the techniques developed and incorporated in the various modules and interfaces, should be applicable in a wide range of application domains and tasks.

In this paper[2] we provide general information about the DenK-project in the present section and about the prototypical DenK-system in the next section, discussing the basic motivation, principles and approaches that underlie the architecture and functionality of the DenK-system and the ensuing research activities. In sections to follow we briefly discuss the main technical work in the project concerned with knowledge representation and reasoning, natural language interpretation, cooperative communication, and domain modelling.

## 2 Theoretical background

The DenK-project takes its starting point in a view on human communication where natural communication is considered as being motivated by some underlying goal, purpose, or function; a view adopted from recent approaches to dialogue analysis, such as *Dynamic Interpretation Theory* (Bunt, 1991; 1994) and *Communicative Activity Analysis* (Allwood, 1995). For human-human communication this underlying motivation may be either of a social character, such as being friendly or polite, or giving or seeking moral support, or courting; or it may concern a concrete task that the agent wants to get done by or with the help of another agent. In the case of human-computer communication, motivations of the former kind do not seem to arise, and we may safely assume that the user is communicating in order to accomplish a certain task - what is often called an *application*. In many kinds of applications, such as process control or computer-aided design, real-world objects or potential real-world objects are involved. Humans naturally interact *physically* with such objects, by means of actions like picking up, turning around, or fastening, and *perceptually*: observing them by seeing, feeling, hearing, or smelling. Interaction with other *agents*, by contrast, is naturally done in a *symbolic* fashion using natural language, gestures, facial expressions, and body language in order to achieve communicative purposes such as giving a command to perform an action on the task domain, or requesting or providing information. The essential difference between the two types of interaction is that symbolic actions (for instance, speech acts) need an interpreter who can bridge the gap between the symbols and their actual meaning and purpose,

---

[2] In this paper we reuse material from Ahn et al. (1995), Beun & Ahn (1995), Bunt & Beun (1992), and Borghuis (1995).

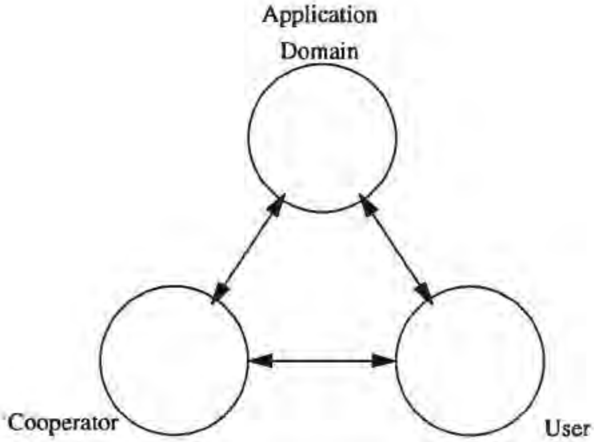


Figure 1: The triangle metaphor: user, cooperator and application domain.

while actions of the second type are related in a more direct manner to human perception and action (Hutchins, 1989; de Souza, 1993).

The two types of interaction are clearly distinguished in the design of the DenK-system, where two components play a crucial role:

**a.** the so-called *cooperator*, who interprets symbolic messages from the user, is capable of reasoning about various aspects of the domain and the user, and produces communicative behaviour adequate with respect to the user's beliefs and goals;

**b.** the *task* or *application domain* model, implemented by means of an animation system which incorporates spatio-temporal components and graphical tools for representation and visualization of the domain.

The user can interact indirectly with the application domain through linguistic communication with the cooperator, who has internal access to the domain, and directly through input and output devices, such as mouse and screen.[3] This is depicted in what we call the *triangle metaphor* of the communication situation (Figure 1), where we have *three* interacting components: the user, the cooperator, and the domain model. We will see below that the overall design of the DenK-system is based on this metaphor.

It may be noted that the overall architecture of the DenK-system as based on the triangle metaphor differs from the architecture of other intelligent multimedia systems, such as AIMI (Burger & Marshall, 1994), WIP (Wahlster et al., 1993), $MMI^2$ (Wilson et al., 1991) and CUBRICON (Neal & Shapiro, 1989), in which there is no direct link between the user and the application domain. An important advantage of our approach is that certain particularly complex aspects of the interaction, such as the visualization of autonomous motion behaviour of objects, do not have to be considered by the cooperator and can be left to the interactive component of the application domain model.

In order to be able to act as the ideal electronic cooperative assistant, the DenK-system should in the first place have a good understanding of what the user wants and does not want, knows and does not know, etc. A good assistant should only need the proverbial half word from

---

[3] And, of course, the user can interact by using these two modalities simultaneously, for instance by using a deictic expression and pointing at an object on the screen.

the user in order to know what to do. Good understanding crucially depends on sufficient and relevant knowledge about each other, and particularly on *shared knowledge.* The DenK-project therefore adopts the approach to communication known as Dynamic Interpretation Theory (DIT, Bunt 1989; 1991; 1994; 1995a; Beun 1994; Bego, 1995), which analyses a dialogue in terms of combinations of actions called *dialogue acts*, defined by the way they change the current dialogue context, notably the current state of knowledge and shared knowledge of the participants. For modelling states of (shared) knowledge, which is a central concern in this approach, the project applies and extends a form of constructive type theory (Ahn, 1992; Helmink & Ahn, 1994; Ahn & Kolb, 1991; Borghuis, 1994). For the visual modelling of the application domain the project follows a succesful line of work in the development of *object-oriented animation* software, which has resulted in a powerful system called the Generalized Display Processor (van Overveld, 1991).

## 2.1 Research areas and approaches in the DenK-project

From the philosophy underlying the DenK-project it follows that the project needs to address research issues related to the design of the cooperator and domain modelling components.

The cooperator, as we have seen, has the task of interpreting the user's natural language messages[4] and deciding what to do on the basis of the interpretation it constructs. In general, the system has to consider the generation of two kinds of actions: task-related actions on the application domain, and communicative actions for continuing the interaction. Both the generation of appropriate actions as well as the contextual interpretation of user inputs require the availability of context information of various kinds, and the ability to reason with that information. The design and development of the cooperator component thus entails research in the following key areas:

- Contextual interpretation of natural language dialogue contributions
- Formal and computational modelling of contextual knowledge
- Automated reasoning facilities, to be effectively applicable during utterance interpretation and output generation
- Formal and computational modelling of communicative actions and their interaction with the context model

According to the triangle metaphor, domain modelling in the DenK-project has two faces, a visual one, on the screen and accessible to the user through a direct manipulation interface; and a system-internal one of a formal symbolic kind. The internal model is needed in order for the cooperator to reason about potential actions on the domain, and it is in general also needed because not all properties of the domain can be visualized (such as constraints on possible objects, and invisible properties such as weight and price). To develop the domain model component, work is thus needed concerning the following issues:

- Formal domain modelling in a way that is computationally feasible and suitable for the reasoning to be done in the cooperator, as well as for the interaction with the visual domain modelling;

---

[4] Possibly augmented with graphical components; see previous footnote.

- Visual domain modelling in such a way that the visual model can effectively exchange information with the formal domain model, as a basis for the interaction with the cooperator and in order to maintain a consistent overall (formal/visual) domain model;
- Supporting a direct manipulation and observation interface with the user, consistent with and supportive for the natural language interface supported by the cooperator.

In order to address these issues, which altogether cover a wide spectrum of interdisciplinary research, the project is divided into a number of subprojects, carried out at the participating locations in Tilburg and Eindhoven. Without being exhaustive, we would like to mention the currently ongoing subprojects plus the most significant ones that have been concluded in the period of 1988 -1994.

1. System design and integration.
2. Type theory and deduction for two-agent belief modelling.
3. The semantics of constructive belief modelling.
4. An object-oriented interface language for visual domain modelling.
5. Empirical investigations into user behaviour in DenK-like situations.
6. Computational pragmatics and dialogue management modelling.
7. Constraint-based grammar and parser development for dialogue utterances in DenK-like situations.
8. Natural language semantics using constructive type theory.
9. Formal modelling of temporal domain aspects in type theory.
10. Constraints in object-oriented animation.

The 10 subprojects cluster around 4 major themes (plus system design and integration), viz. *knowledge representation and reasoning in type theory; domain modelling through object-oriented animation; context-sensitive natural language interpretation,* and *computational processing of dialogue acts.* We briefly consider each of these topics and relate them to the subprojects concerned. Subsequent sections will be devoted to the research concerning each of these topics, after we have discussed the system design in the next section.

### Knowledge representation and reasoning in type theory

The subprojects 2 and 3, completed in 1993 and 1994, respectively, have both contributed to this topic;[5] project 9 currently extends this work.

As mentioned above and explained in more detail below, we have chosen to use the powerful logical formalism of *type theory* to model states of (individual and shared) knowledge in the DenK-system. Research in the DenK-project in recent years has shown that the logical constructs of type theory called *contexts* (!) are very attractive for this purpose (Borghuis,

---

[5] Both subprojects have resulted in published PhD dissertations: Jaspars (1993) and Borghuis (1994), respectively.

1994; Ahn, 1992; Jaspars, 1993). The expressive capabilities and proof methods of standard type theory have to be enriched, however, for adequate modelling of the states of information and intention of agents participating in a dialogue. An extension with epistemic modalities has been defined by Borghuis (1994). This extension allows the formal modelling of beliefs with different degrees of certainty, and of distinguishing 'private' beliefs from assumed shared beliefs (or 'mutual beliefs'). Further extensions are required for the type-theoretical representation of the time-dependent aspects of the behaviour of objects in the application domain, and also for the representation of and the reasoning about the temporal aspects of natural language utterances.

### Domain modelling through object-oriented animation

The DenK-project has chosen for graphical representation of visualizable domain aspects on the basis of object-oriented descriptions. This requires an interactive animation system, since changes in the domain during the dialogue have to be visualized, and the visualized objects should be available for direct manipulation by the user. Such an interactive system has been developed in recent years at Eindhoven University of Technology, the GDP (Generalized Display Processor) (van Overveld, 1991). Within the DenK-project, an object-oriented interface language to the GDP has been developed in subproject 4, called LOOKS (Language for Object-Oriented Kinematic Specification; Peeters, 1994). Given a description in LOOKS of the objects defining the current state of the domain, a 3D graphical respresentation is generated. This animation is indeed interactive, which means that the GDP allows the LOOKS description to change during the animation process. A required extension of the currently available system is that constraints, specifying invariants of the domain, can be introduced (such as impenetrability of objects); subproject 10 is currently addressing this issue.

### Context-sensitive natural language interpretation

The interpretation of the user's dialogue utterances, which is the theme of the subprojects 6 and 7 (with empirical support form subproject 5), should result in a formal representation of the current goals and state of information of the user, given the preceding dialogue and the preceding direct interaction between the user and the domain model. The interpretation of the user's utterance should thus fit into the context created by the preceeding interaction between user and system. As such, the interpretation process is highly dependent on that context. According to the approach of Dynamic Interpretation Theory, which has been adopted in the DenK-project, the interpretation process should deliver a formal description of how the current context is to be updated; for the representation of these contexts we have chosen to use the formal structures of type theory known as (type-theoretical) 'contexts', which are somewhat comparable to the representation structures of Discourse Representation Theory (Kamp & Reyle, 1993). The work on this aspect of the DenK-project has started only recently and is still in an early stage.

### Computational processing of dialogue acts

The effect of the interpretation of the user's dialogue contributions is, as we have seen, that the formal model of the current dialogue context, notably of the state of information and current goals, is created and updated cumulatively during the dialogue. The way in which an utterance affects the current context depends not only on its semantic content, but also on its communicative functions. The research on this subject, which is undertaken in the subprojects 8 and 1, aims primarily at the formulation of systems of rules which, given an interpretation of a natural language input as a dialogue act with a certain semantic content and communicative functions, specify precisely how the the current context model is to be updated, and what kind of dialogue acts are appropriate for continuing the dialogue, given the updated context model.

## 3 The DenK-system

As mentioned above, the DenK-system is intended to be *generic* in that its architecture, as well as many of the techniques developedn and incorporated in the various modules and interfaces, should be applicable in a wide range of application domains and tasks. To demonstrate the generic character of the system, several applications are envisaged, the most important one, to be developed first, being a training simulator for the use of a modern electron microscope as developed by Philips Electron Optics. This device is an important instrument for materials research in physics, as well as for medical research in pathology. The device is very complex, and uninitiated users have to go through an intensive training period in order to learn how to use the device. This is primarily due to the fact that users find it difficult to form an adequate picture of the internal workings of the device (the relevant parts, their functions, and their relations), as the device manifests itself to users essentially as a black box. Until now, the only way to learn how to use the device is to be trained explicitly in its use. Philips expects that the training period can be considerably shortened if the trainee can exercise with an interactive training simulator. This particular application is currently under development in the project.

For experimental purposes, in 1993-'94 a partial prototype of the DenK-system has been built using a toy world of blocks of different shapes, sizes and colours, with speciable autonomous movements in 3D space such as rotating, moving to or from a distance, etc. This blocks world has a 3D-representation on the screen. This partial prototype does not allow direct manipulation of objects by the user, and supports very simple dialogue behaviour in a very small subset of natural language, being able to execute simple commands to act on the domain and to answer questions about its current state.

The supported dialogue behaviour with the system at this point is still very primitive, but it does allow for example to detect presupposition violations. If the user gives a command that is impossible to perform on the domain, the system will report this. Another interesting point is that, in answering questions, the system takes the preceding dialogue into account and only supplies information which, according to its information state, are not recorded as shared knowledge of user and system. The interactive behaviour of this preliminary system is primitive, primarily as a result of the fact that the project work on interpretation of natural language utterances was still in an early stage; it was decided in the project to follow an 'inside-out' strategy, concentrating the work in the period 1989-1994 on first developing the formalisms and the techniques for knowledge representation, reasoning and domain modelling. This has resulted in the availability in the preliminary prototype of the visual domain model component and in the type-theoretical representation-, reasoning- and evaluation systems that form the backbone of this partial prototype.

The architecture of the first (partial) prototype is displayed in Fig. 2 below. The following ingredients make up this system:

- A generic domain emulator for a wide variety of domains (GDP, Generalized Display Processor), capable of simulating a domain given a description a LOOKS description;

- An implemented model of the blocks world;
- A proof checker for type theory, incorporated in the inference module called HOLMES, based on theoretical work in type theory (see e.g. Barendregt, 1991), which checks whether a type-theoretical context is well-formed, in particular whether all expressions in the context are correctly typed.
- A representation system for states of information related to different epistemic modalities and two agents, based on theoretical work using nesting of type-theoretical contexts by Borghuis (1994).
- A reasoning system for type theory (another part of the HOLMES module), based on Helmink & Ahn (1991).
- An internal interface between the cooperator and the domain emulator, called 'DABAS'.
- An evaluator for converting an expression in type theory into a first-order intermediate language and, via the DABAS interface, checking against the current state of the domain. This involves translating type-theoretical expressions into combinations of domain primitives, known to the domain model.
- An implemented set of (preliminary, highly simplified) rules for updating type-theoretical representations of the system's information state, given an interpretation of an input dialogue utterance, and for generating appropiate reactions in terms of dialogue acts.
- A toy parser for natural language, with an implemented miniature fragment of Dutch.[6]
- An implemented preliminary and highly simplified version of the intermediate language ULF (Underspecified Logical Form, Kievit, 1994), which makes the connection between the parser and the HOLMES module. This language is based on previous work in the Core Language Engine project (Alshawi, 1992), in the Esprit project PLUS (Geurts & Rentier, 1993) and in the ΔELTA project (Rentier, 1993; Bunt, 1995b).

In the second half of the project (1994-1998), the emphasis will be on the devlopment and implementation of the techniques for natural language interpretation and dialogue management and on specifying and developing the components necessary to deal adequately with the electron microscope training simulation application. In addition, many of the components of the first partial prototype are of the character of provisional, experimental implemntations, and will be redeveloped systematically for the definitive prototype system.

## 4 Knowledge representation and reasoning

### 4.1 Type theory

To model the information state of the cooperator and particular aspects of the communication process, we use a versatile and powerful formalism called *type theory*. Type theory refers to a class of formalisms, including *Automath* (de Bruijn, 1980), *Intuitionistic type theory* (Martin-Löf, 1984) and the *Calculus of Constructions* (Coquand, 1985), which are all based on similar

---

[6] For the protoype system to be completed at the end of the project, the envisaged natural language is English.

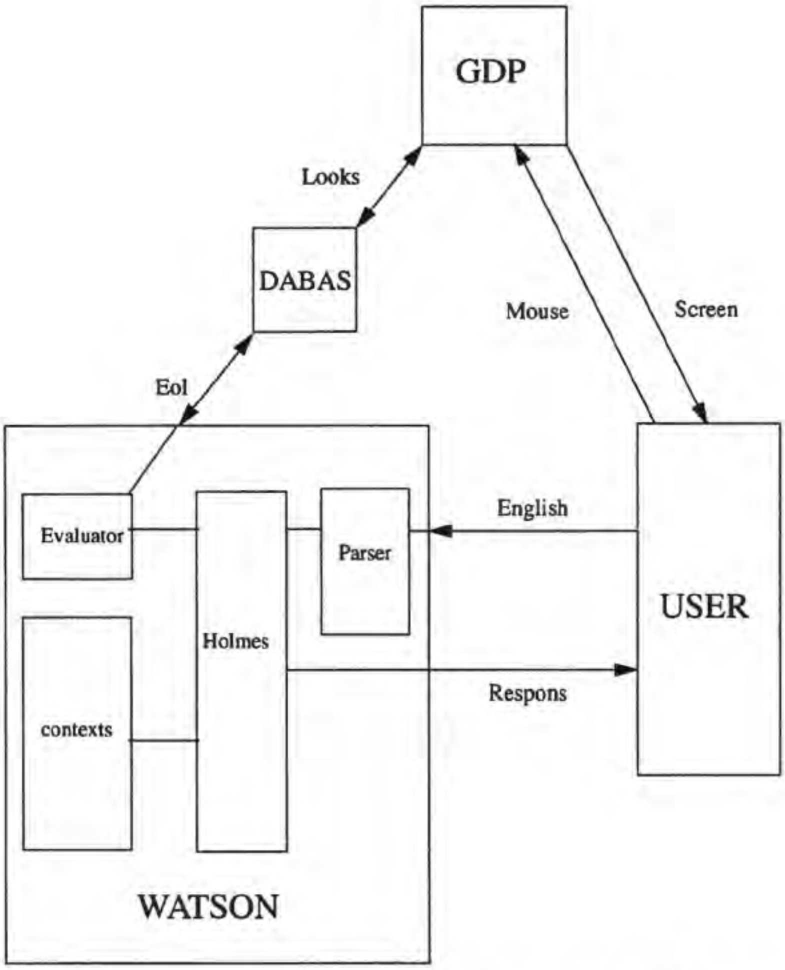


Figure 2: Architecture of the first partial prototype of the DenK-system.

ideas (see also Barendregt, 1991). These formalisms function as logical frameworks in which almost any form of mathematical reasoning can be expressed. They find their origin in foundational mathematical research, and have a strong constructivist flavour.

We will use type theory not only as a formalism to model the cooperator's beliefs, but also for the semantic representation of the natural language utterances exchanged between the cooperator and the user. Mäenpää & Ranta (1990) have already pointed out that type theory may be very useful for this purpose. Among the many advantages of type theory over other formalisms we would like to mention in particular the inherent *dynamics* of the formalism, and the built-in notion of the *justification* of propositional information (cf. van Benthem, 1991).

**Dynamics:** The representation structures of type theory, as already mentioned, are sequential structure that can be extended in different ways to incorporate new information. The ways in which these structures, called *contexts*, may be extended to accommodate new information, are laid down in so-called *introduction rules*. In addition, *inference rules* describe how information may be combined in a sound way so that implicit information within a context can be made explicit. In a similar way, objects of various kinds are constantly introduced and added to the structure that grows as a dialogue proceeds (see below for a worked example). This is similar in spirit to what happens in Discourse Representation Theory when a representation structure is built incrementally as the analysis of a discourse fragment proceeds. In fact, it has been shown in Ahn & Kolb (1991) that type-theoretical contexts can be regarded as natural generalizations of Discourse Representation Structures (Kamp, 1981).

**Justifications:** Another important notion in type theory is that of a *proof*. Proofs are considered as mental objects, just like individual concepts, have concrete representations, and are completely integrated within the formalism. This means that within type theory we can not only represent *what* the cooperator believes, but also *how* he comes to believe it, by having explicit representations of the proofs that justify his beliefs. Type theory is not so much concerned with truth, however. It records information in (type-theoretical) contexts and what has been shown to follow from this information. The theory is *constructivist* in that it considers propositions as being proven only if a proof has actually been constructed and is explicitly present in the (type-theoretical) context. This also means that partiality is inherent to the formalism.

### 4.2 Information states as type-theoretical contexts

In this section we show, by means of an example, how the cognitive state of the cooperator can be represented by a type-theoretical context.

We consider a simple blocks world example where we have pyramids, cubes and the like, which can move around and have properties like colour and size. In this example we assume the cooperator to have the following information:

- there may be pyramids and colours in the domain (the cooperator is familiar with the notions 'pyramid' and 'colour'; see below);
- there is actually at least one pyramid in particular ;
- 'small' is a predicate on pyramids, and 'bright' is a predicate on colours;

- all pyramids have a colour, and every small pyramid has a bright colour.

A type-theoretical context is a sequence of assignments of *types* to objects. Objects and types are denoted by expressions, with the infix operator ':' used to relate objects and their types; 'O : T' should be read: 'O is an inhabitant of T', or 'O is an object of type T'. Expressions of this form are called 'entries'. The entries making up a context are separated by commas; beginning and end of a context are marked by '[' and ']', respectively.

To describe the cooperator's information state, we begin with an empty context and gradually introduce the information listed above. First, we have to introduce the necessary types corresponding to the concepts whose familiarity is assumed, i.e. 'pyramid' and 'colour'. In type theory, new types may be introduced whenever they are needed as inhabitants of the *supertype*, denoted by '$*$', which is the mother of all types.[7] Accordingly, we obtain a context with the following two entries:

[ pyramid : $*$ , colour : $*$ ]

After a type has been introduced, one may introduce inhabitants of that type. In this case we introduce an anchored specimen of a pyramid, by giving it a name and extending the context with the entry:

p318 : pyramid

The introduced pyramid is labelled by the expression 'p318' (all introduced objects must be uniquely identifiable).

In type theory, predicates are represented as functions to propositions. Viewed in this way, the predicate 'bright' is a function that, given a particular colour, yields a proposition like: 'Yellow is bright'. In type theory, *propositions are treated as types*, i.e. they occur at the same level as the types 'pyramid' and 'colour' in the present example; propositions themselves have type '$*$'. An inhabitant of a proposition (type) is viewed as representing a proof of that proposition. Proofs and other objects may be combined in order to construct new proofs and other objects. The inference rules of the type system restrict the way in which this can be done, and guarantee that all reasoning is sound.[8]

Given this interpretation of propositions, we can introduce the predicate 'small' as a function from pyramids to propositions, and the predicate 'bright' as a function from colours to propositions (objects of type '$*$'). These are added to the example context by extending it with the entries:

small(M) : $* \Leftarrow$ [ M : pyramid ]
bright(C) : $* \Leftarrow$ [ C : colour ]

These entries need some clarification, since their right-hand sides, representing function types, are not of the general form 'O : T'. It is possible in the general form, but to show this would require a rather long and formal exposition. Instead, we adopt a simplified Prolog-like notation

---

[7] Except of itself, of course, in order to avoid logical paradoxes. The existence of this supertype, which functions as the 'type of the types'is implicitly postulated by type theory, in the sense that it may be used in every context without being introduced in that context.

[8] This ingenious idea is known as the Curry-Howard isomorphism (Curry & Feys, 1958).

in which a limited fragment of type theory can be expressed in a rather intuitive way. The symbol '$\Leftarrow$' in this notation corresponds to the Prolog turnstile ':-'. The two functions in our example are expressed as Prolog-like clauses in this notation, which can be read as: "if M is an object of type 'pyramid', then the function 'small', applied to M, yields a proposition.

Using the clause notation and the predicates, we can express the cooperator's knowledge that pyramids have a colour, and that small pyramids have bright colours. We first extend the context with an entry representing a function, 'c', that associates colours to pyramids:

$$\mathrm{c(M) : colour} \Leftarrow [\ \mathrm{M : pyramid}\ ]$$

Finally, we have to express that every small pyramid has a bright colour. To this end, we introduce a function (called 'ax2') that, (the object 'P' of type 'small(M)'), returns a proof that the colour of this particular pyramid is bright (the object ('ax2(M,P)') .

$$\mathrm{ax2(M,P) : bright(c(M))} \ \Leftarrow [\ \mathrm{M : pyramid}\ ,\ \mathrm{P : small(M)}\ ]$$

Combining all the above entries results in the following context, which represents the beliefs of the cooperator about the domain:

$$\begin{array}{l}
[\ \mathrm{pyramid} : * ,\\
\quad \mathrm{colour} : * ,\\
\quad \mathrm{p318 : pyramid} ,\\
\quad \mathrm{small(M)} : * \ \Leftarrow [\ \mathrm{M : pyramid}\ ] ,\\
\quad \mathrm{bright(C)} : * \ \Leftarrow [\ \mathrm{C : colour}\ ] ,\\
\quad \mathrm{c(M) : colour} \ \Leftarrow [\ \mathrm{M : pyramid}\ ] ,\\
\quad \mathrm{ax2(M,P) : bright(c(M))} \ \Leftarrow [\ \mathrm{M : pyramid}\ ,\ \mathrm{P : small(M)}\ ]\\
\quad ]
\end{array}$$

Once the information is represented in the form of a type-theoretical context, the cooperator can make inferences by constructing new objects using the entries in the context. In fact one can extend the Prolog-like notation used here in such a way that it supports an effective proof construction method for type theory, combining resolution style proofs with natural deduction (Helmink & Ahn, 1991). The cooperator contains a theorem prover based on this

## 4.3 Communication in type theory

The cooperator's is primarily to communicate with the user, i.e. to interpret the user's utterances and to generate appropriate responses.[9] To achieve this, the cooperator uses the following three main sources of information

- the application domain, as modelled visually;
- the private beliefs, as represented in type theory.
- the assumed mutual beliefs, as represented in type theory;[10]

---

[9] As noted above, the current implementation has no provisions for natural language output; all reponses are provided in type-theoretical formulas.

[10] Note that besides the mutual beliefs, the cooperator also shares the visual domain model source with the user: the user has direct visual access to this model.

In the current DenK-system, the cooperator distinguishes between the kinds of belief by maintaining *two* contexts: a *private* context representing his own beliefs, and a *common* context representing the mutual beliefs. Recently, an extension of type theory has been developed which allows representation of and reasoning about different kinds of belief in one context (Borghuis, 1993; 1994).

We present two cases in which the cooperator generates a pragmatically correct reply to an utterance of the user, using the information sources. In the first case the user makes a statement, in the second he asks a question. In both cases, we assume that the information state of the cooperator is the one represented above, and that all beliefs of the cooperator are mutual beliefs except for the constraint that every small pyramid has a bright colour, which is a private belief of the cooperator. Remember that we assumed the cooperator to be an expert about the domain and that all declaratives about the domain contributed by the user are interpreted as questions.

Having inspected the image on the screen, the user might produce the following sentence:

"The pyramid is small"

In order to interpret this utterance the cooperator has to figure out which pyramid is meant. Due to the definiteness of the noun phrase, the cooperator may consult both the application domain and the mutual beliefs. The cooperator's private beliefs are irrelevant here, because the user is unaware of those, and hence cannot take them into account to produce a definite reference. The cooperator will therefore assume that the user refers to the pyramid about which a mutual belief is stored in his common context ('p318 : pyramid'). He will interpret the user's utterance as saying that there is evidence that 'small(p318)' holds. After checking the truth of this proposition in the application domain, the cooperator extends his *common* context with the following entry;

e435 : small(p318)

where 'e435' labels the evidence the cooperator has found for the proposition 'small(318)'. The cooperator will give the answer "yes", to indicate that he agrees. The proposition **the pyramid is small** becomes a mutual belief of cooperator and user.

Suppose that the next utterance of the user is:

"Is the colour of the pyramid bright?"

The communicative function of the utterance is a **yes/no-question**; the user signals that he wants to know whether the proposition **the pyramid is bright** holds. The cooperator complies with this wish by trying to construct an object for the proposition ('bright(c(p318)'). Using the theorem prover, the cooperator succeeds in constructing such an object from the entries in his private context:

ax2(p318,e345) : bright(c(p318))

Because the question was a **yes/no-question**, only the existence of a proof-object matters. The cooperator has found a proof-object, hence the question will be answered affirmatively.

If the user had asked:

"Why is the colour of the pyramid bright?"

the cooperator would be under the obligation to communicate how he came to believe 'bright(p318)'. This is recorded in the proof-object 'ax2(p318,e345)', but by the Gricean maxim of quantity the cooperator should only communicate those ingredients of the proof that are not already among the beliefs of the user. By checking the common context, the cooperator can find out that the user already believes that there exists a pyramid ('p318 : pyramid'), and that it is small ('e345 : small(p318)'. Hence, the cooperator generates an answer from the only ingredient in his proof that is not in the common context:

"Because every small pyramid has a bright colour".

This answer is satisfactory since it provides the user with only the *new* information needed to infer how the cooperator came to believe that the pyramid was bright.

In answering WH-questions, other complications also occur. It may be particularly difficult to communicate the identity of an object, even if it occurs in the common context, because the message to the user should use the properties of the object to find a description that identifies the object unambiguously for the user. Which description is actually most appropriate is a complicated matter, depending on aspects such as the difference in salience of the properties of the objects, previous utterances, and domain focus (see Cremers, 1994; 1995). We have not yet implemented the generation of such answers; in the current implementation the cooperator will simply 'point out' the desired object by highlighting it in the graphical domain representation on the screen.

## 5 Natural language interpretation and dialogue management

The view underlying the design of the cooperator in the DenK-architecture, which is responsible for the analysis of the user's dialogue contributions and the decision of what actions (domain actions and communicative actions) to perform in view of the input analysis, is that dialogue participants use language to perform communicative acts, primarily aimed at changing the addressee's cognitive state in the direction of the speaker's goals (Bunt, 1989; Bunt, 1993). In line with Searle (1969), two aspects are distinguished in the communicative act: its *semantic content* and its illocutionary force, or what we will call its *communicative function*. The semantic content is related to the truth-conditional aspects of the action, such as the existence of particular objects, their properties and relations in the application domain. The communicative function determines, together with the semantic content, the effects of the communicative act on the cooperator's information state.

### 5.1 Linguistic analysis

The user of the DenK-sytem should be able to communicate with the system by means of the keyboard, using expressions of English (with syntactic, lexical and other limitations). The system should recognize the communicative functions and semantic contents of the user's utterances, en update its information state accordingly.

In linguistic analysis, three aspects are traditionally distinguished: (morpho)syntactic, semantic and pragmatic analysis. Syntactic analysis is concerned with parsing a complex expression into its constituent parts; semantic analysis is concerned with the semantic consequences of the syntactic analysis as well as with the interpretation of lexical items, and pragmatic analysis ties semantic analysis to aspects of the context of use. In the DenK-project we have chosen to use a modern framework for linguistic analysis which emphasizes the *integration* rather than the *separation* of these aspects of analysis, as has been customary until recently. This framework, called *Head-driven Phrase Structure Grammar* (HPSG), does not build up syntactic tree structures, as other grammatical formalisms usually do, but produces complex *typed feature matrices* which incorporate both syntactic, semantic, and pragmatic information. These feature structures can be viewed as representations of partial information states (Pollard & Sag, 1987), and are as such attractive for the DenK-project where partiality of information states is a pervading phenomenon.

After considering and evaluating a number of alternatives, we have decided to use to publicly available *Attribute Logic Engine* (Carpenter, 1994) for HPSG-based linguistic analysis. However, where HPSG standardly comes with a semantic interpretation that is not very suitable for the DenK-project, we have decided to add to HPSG and ALE a different interpretation component that is better suited for context-sensitive interpretation and the use of type theory for representing information states. This interpretation component makes use of an intermediate level of expressions, mediating between feature matrices and type-theoretical contexts. For this intermediate level we use a representation language called 'ULF', for '*Underspecified Logical Form*', which allows semantic representations that are 'underspecified' in the sense that they may leave open a variety of aspects of the semantics of the natural language expression under consideration, such as the relative scopes of scope-bearing elements, the logical interpretation of natural language quantifiers, or the interpretation of anaphoric pronouns. The ULF language, of which a preliminary design has been provided by Kievit (1994), is based on the Quasi-Logical Form language of the Core Language Engine (Alshawi, 1992), and on more recent work in the projects PLUS (Geurts & Rentier, 1993) and ΔELTA (see Rentier, 1993; Bunt, 1995b).

As far as pragmatic analysis in the cooperator components is concerned, the pragmatically relevant aspects of natural language inputs, as recorded in feature matrices, are extracted and stored in a list of pragmatic attributes and their values. These attribute-value lists are interpreted as communicative functions, using pragmatic interpretation rules that we consider briefly below.

It should be emphasized that the DenK-research concerned with linguistic analysis and dialogue pragmatics is currently in a relatively early stage, as already noted above; this work is occupying a central position in the project in its second four-year period (1994-1998).

## 5.2 Pragmatics and dialogue management

The cooperator's communicative) behaviour is controlled by the *pragmatic rule interpreter* (see Figure 3). To produce simple cooperative behaviour, the interpreter exploits three types of contextual information:

1. the information state of the cooperator;
2. the most recent (communicative) action performed by the user;

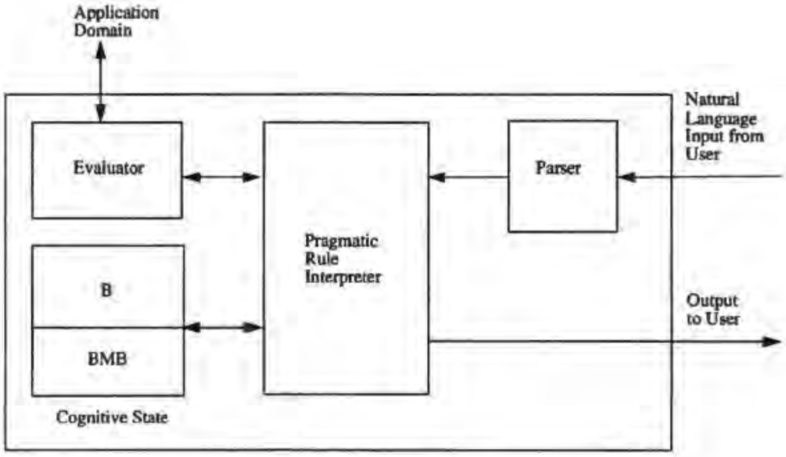


Figure 3: The design of the cooperator.

3. the current state of the application domain.

First, the interpreter analyses the user's utterance within the current context and, if no communication failures are noted, it updates the information state of the cooperator with the new information. If the communicative function of the user's utterance was a command, a domain-related action is performed; if the function was a question, the question is answered or appropriately responded to if no straight answer can be given. Since we consider the cooperator as an expert about the domain, he does not accept new domain information from the user, and consequently, declarative statements by the user about the domain are interpreted as questions (see Beun, 1989).

One of the things the cooperator has to do as part of its contextual interpretation of utterances, is linking objects that are mentioned in an utterance of the user to entities in the application domain. In the case of a definite reference, the cooperator looks for a specific red block that has been introduced in the dialogue before and therefore belongs to the cooperator's assumed mutual beliefs, or that can be detected in an unambiguous way in the current state of the discourse domain by inspecting its visual representation. In the case of an indirect reference, as in "Move a red block", no specific object needs to be identified, and the choice of the object is left to the cooperator.

New variables that arise from the introduction of definite and indefinite objects in the discourse are linked to entities in the domain by means of so-called *satisfying assignments* (Ahn & Kolb, 1991). For instance, the variable that results from interpreting "the red block'" has to be linked to a suitable object in the cooperator's mutual beliefs; if the link can be established, the user can refer anaphorically to the object in subsequent utterances. On the other hand, if no such link can be established (for instance, because there is no such object), the cooperator generates adequate feedback to signal this.

Having extended the current information state with the new information and linking the introduced objects to domain entities, the cooperator has to generate an adequate reaction. For that purpose, we consider the user's goals as they follow from the analysis of the semantic content and the communicative functions of the users's utterance. In the implemented partial prototype we have assumed for the time being a simple and straightforward relation between the an utterance's communicative functions and the underlying user goals: commands are used to *change* the state of the application domain and questions are used to *obtain* certain information about the domain. In the second phase of the DenK-project we will incorporate more sophisticated pragmatic rules that take multifunctionality, indirectness, and dialogue control mechanisms into account (cf. Beun, 1991; Bunt, 1994; 1995a).

As already mentioned, the cooperator should consider two types of action in response to the user's communicative behaviour:

- *domain acts* that are directed towards a change in the state of the domain;
- *communicative acts* that are directed towards a change in the information state of the user.

Domain acts are generated in reaction to a command of the user. In those cases, the cooperator has to find the procedure that corresponds to the given command and that can be executed by the domain application.

The interpretation process can only be executed successfully if certain conditions are fulfilled. For instance, all content words in the user's utterances should be interpretable in terms of elements in that part of the cooperator's information state that represents mutual beliefs. If the user makes assumptions that are not part of the cooperator's belief (private or mutual), these will be corrected by the cooperator, by means of *corrective* dialogue acts. For instance, to answer the question "Why is the red block rotating?" the cooperator should look for a uniquely identifiable red block in the mutual beliefs or in the visual domain model before the answer can be provided. If no such object is available, a correction towards the user is performed.

If the interpretation process is executed successfully, this means in practice that the cooperator can try to provide an answer to a question of the user, or to execute certain domain actions. We have already seen, in the example concerning the the user of type theory, how the use of mutual beliefs is crucial for generating appropriate answers to the user's questions, as it is the basis for providing accurate explanations (answers to why-questions) as well as for making a good choice of properties to identify objects in answering WH-questions.

## 6 Visual domain modelling: the Generalized Display Processor

### 6.1 DenK-requirements

The predominant requirement for the visual domain representation module is that it can represent the essential visual (spatial) and dynamic (temporal) aspects of this domain in such a way that a convincing real-time view is presented to the user. This can be achieved by using an animation system, provided that it meets the following requirements:

- in order to support multi-modal interaction:
  1. interrogation of the domain status has to be allowed at any time, i.e. asynchronously with respect to the time evolution of the animation;
  2. the display image has to be refreshed continuously, so that the user rceives visual feedback of the animation irrespective of the state of the dialogue;

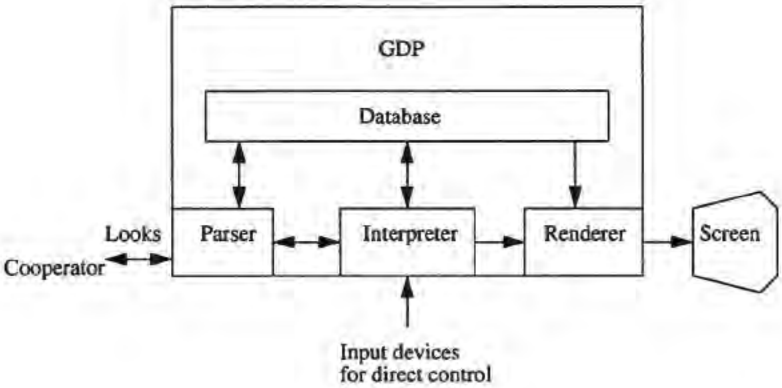


Figure 4: The Generalized Display Processor.

3. the user has to be able, at any time, to refer to aspects of the graphical model both via the cooperator and via direct manipulation, using a mouse and a 3D-simulated pointing device.
4. in order to support a generic platform to simulate a large variety of application domains:
    (a) there should be instructions to create objects[11] (both their geometrical shape and their autonomous motion behaviour) and to pass messages to objects to alter their properties and behaviour; objects also have to be able to pass messages to each other;
    (b) to facilitate programming complex behaviour (e.g., the motions of mechanical devices, walking, grasping objects) a library of versatile built-in motion methods has to be available.

In order to meet these requirements, we have developed an architecture called the (Generalized Display Processor (GDP, van Overveld, 1991); see Figure 4). The GDP is a virtual processor consisting of a parser, an interpreter, rendering support, and a database.

The parser provides for the communication with the cooperator; it checks the incoming information and, if necessary, sends messages back from the interpreter to the cooperator. The interpreter calculates the values of the attributes of the objects (e.g. position, speed) in order to generate a new frame of animation, which is displayed on screen by means of the renderer[12]. The behaviour of the objects may be autonomous, or dictated by the cooperator or by input devices for direct control (e.g. the mouse).

At any discrete time, a complete description of all geometrical attributes of a moving scene is stored in the database. The database consists of two tables: the class table and the object table. The class table contains the class definitions that state which attributes exist

---

[11] Here, we use terminology from object-oriented programming. An 'object' is a variable containing both data (its 'attributes') and optionally some program fragments (its 'methods'). Calling a method of an object is referred to as passing a 'message' to that object. An object is an instantiation of a 'class'; the class definition lists all available attributes and methods for all objects to be instantiated programming, we refer to Meyer (1988).

[12] Provided the time granularity of the animation is sufficiently small and the renderer is sufficiently fast, the impression of a moving display results.

for a particular class as well as which methods it may execute.[13] The object table contains the actual objects in the application domain, i.e. instantiations of existing classes.

### 6.2 The control language LOOKS

The language LOOKS (Language for Object Oriented Kinematic Simulations) has been defined for programming the GDP (i.e. in order to define what classes of objects will inhabit the application domain, to create these objects, to program the motion methods, to pass messages and to execute several other types of statements) (Peeters, 1994).

LOOKS supports a variety of object-oriented features, such as data hiding, abstract data types, strong typing, genericity and multiple repeated inheritance; it implements (quasi-)parallellism to facilitate the specification of concurrent motions.

At any time, fragments of LOOKS texts may be passed to the parser. Such a fragment can be either a class definition, an object definition or a message. If this fragment is successfully parsed, and if it was a

*class definition,* then an entry is created in the class table. Apart from the user-defined classes, LOOKS supports a variety of pre-defined classes, including integers, reals, vectors, movable geometric

*object definition,* then an entry is created in the object table;

*message to an object,* then the corresponding method is handed to the interpreter to be executed.

Since the cooperator interacts via LOOKS with the GDP (see Figure 4), there is a close relation between LOOKS and the type theory formalism that was introduced previously in this paper. Virtually, type-theoretical expressions are semantically grounded in LOOKS and the translation procedure comes down to a standard evaluation process[14] where complex type-theoretical expressions are expanded into basic expressions that can be interpreted in LOOKS (Ahn, 1994).

### 6.3 Frame generation

In order to generate a frame of animation, the interpreter installs a method it receives from the parser into the list of active methods and executes the methods it encounters. Most active methods can be fully executed (e.g. assignments, expression evaluations, object transformations), after which they are removed from the list.

Methods may, however, also contain a 'synchronize' statement (typically within the body of an (infinite) loop). Execution of an active method proceeds until it is either fully executed, or until a 'synchronize' is encountered. In the latter case, execution stops, but the method is kept in the list; it will again receive the interpreter's attention when preparing the next frame. Since several methods may contain a 'synchronize', several of which may be kept in the active method list simultaneously, the synchronize mechanism may serve to achieve quasi-parallellism.

---

[13] A class may be defined from 'scratch' or it may inherit attributes and/or methods from other classes. The types of its attributes, as well as the types of the formal parameters in its method headings and the return types of its methods may either be earlier defined classes or generic types.

[14] See also the 'evaluator' in Figure 3.

Also, the execution of a method, due to passing an asynchronous message, may invoke the execution of another method by calling this method. The latter method is also put into the list of active methods, and it will be executed as well during the preparation of the frame.

If no more active methods can proceed any further, the preparation of the frame is complete, and a snapshot of the objects is rendered making use of the active light sources and simulated material properties (colour, shininess, etc.).

Part of the preparation of each frame is also the taking into account of user interaction. Via mouse events and appropriate LOOKS system methods, the user may interact with the ongoing animation (e.g. selecting an object and calling the mouse notifier method that has been assigned to this object).

The current GDP-implementation is designed to run on a high-end graphics workstation where it produces a flicker-free display of shaded images, illuminated by simulated light sources, consisting of several hundreds of polygons with a frame update rate between 15 and 20 frames/sec. It covers most of the requirements and functionality as described above (direct manipulation interaction tools and versatile built-in motion methods are presently under construction).

## 7 Conclusions and future work

In the first four years of the program (1989-1993), most of the effort has been spent on formal aspects of communication, such as formal context modelling, representation of the beliefs of the dialogue participants, and rule-driven generation of dialogue acts. A first, provisional prototype of the DenK-system has been built and its architecture reflects the above view on the conceptual relations between user, cooperator and application.

A central issue in our approach is to develop an architecture for user interfaces that enables us to formulate and to implement rules for cooperative behaviour, independent of a particular application domain. The design of the interface does not originate from the desire to model particular natural language phenomena, but from the need to establish *natural communication* between the application domain and its user, independently of the surface structure of the message. It is our opinion that phenomena well known in natural language semantics and pragmatics – such as context-dependency of the message, Gricean maximes, (in)definite reference, deixis – follow naturally from the fundamental properties of communication, even within the relatively simple model that has been presented in this paper.

In the next four years we will focus on the development of the natural language component and the refinement of the information state modelling in the cooperator. We will study the extension of type theory with temporal aspects and modalities that are essential for describing adequate communicative behaviour, such as different types of belief and intentions. We have planned to develop and evaluate different rules for cooperative behaviour, based on the notions that were introduced in this paper and supported by experimental work in dialogue research. Finally, we will move to the realistic domain of the electron microscope training simulation, and incorporate a so-called 'constraint-specification mechanism' that describes the properties of the domain that remain constant during the interaction. Of course, in the future the cooperator should be able to reason about these constraints.

## Acknowledgements

We would like to thank the organization for inter-university cooperation between the universities of Tilburg and Eindhoven ("Samenwerkingsorgaan Brabantse Universiteiten") for providing most of the funds, and secretary-general Marianne Wagemans and her predecessor Alice Pillot for providing logistic support to make the DenK-project possible. The research reported here has been or is carried out in the Institute for Language Technology and Artificial Intelligence (ITK) and the Research Unit for Language, Information, and Knowledge Engineering (TIK) at Tilburg University; in the Department of Mathematics and Computer Science at Eindhoven University of Technology, and at the Institute for Perception Research IPO in Eindhoven, a joint venture of Philips Research and Eindhoven University of Technology. We owe thanks to many people at these sites for providing the infrastructure and all kinds of support in carrying out the research; special thanks are due to those who provide supervision at the various sites, notably Rob Nederpelt, Kees van Overveld and Jos Baeten at Eindhoven University of Technology, Harry Bunt and Reinhard Muskens at Tilburg University, and Robbert-Jan Beun and Don Bouwhuis at IPO. Harry Bunt at Tilburg University is responsible for the overall supervision of the DenK-project. A word of special thanks goes to Hugo van Leeuwen from Philips Electron Optics, for most valuable support in developing the training simulation application of the DenK-system.

The following people, besides the authors and supervisors mentioned above, have contributed to or are involved in the DenK-project: Gino van den Bergen, Anita Cremers, Frens Dols, Jan Jaspars, Leen Kievit, Paul Piwek, Eric Peeters, Gerrit Rentier, Rita Roelfs, and Margriet Verlinden.

## References

AHN, R.M.C. (1994) *The Database.* Internal Report, Tilburg University. (to appear)

AHN, R.M.C. & KOLB, H.P. (1990) Discourse Representation meets Constructive Mathematics. In: Kálmán, L. & Pólos, L. (eds.) *Papers from the Second Symposium on Logic and Language*, Akadémia Kiadó, Budapest. 105-124.

AHN, R., KIEVIT, L., RENTIER, M. VERLINDEN, M. (1995) The DenK-approach: multi-level semantics in dialogue. To appear in *Proc. CLIN'94.*

AHN, R. (1992) A type-theoretical approach to communication. *THINK Quarterly 1(1).*

AHN, R.M.C., BEUN, R.J., BORGHUIS, T., BUNT, H.C. & VAN OVERVELD, C.W.A.M. (1995) The DenK-architecture: a fundamental approach to user interfaces. *Artificial Intelligence Review 8 (2).*

ALLWOOD, J. (1995) Pragmatics and communicative action. In: Black, W.J. & Bunt, H.C. (eds.) *Abduction, Belief and Context*, London University Press. (to appear)

ALSHAWI, H. (1992) *The Core Language Engine*, Cambridge: Cambridge University press.

BARENDREGT, H. (1991) Introduction to Generalized Type Systems. *Journal of Functional Programming, 1(2).* 125-154.

BEGO, H. (1995) Context Change and Communicactive Feedback. In Beun, R.J., Baker, M. and Reiner, M. (eds.) Dialogue and Instruction: Modelling Interaction in Intelligent Tutoring Systems. Berlin: Springer Verlag. (to appear)

BEUN, R.J. (1989) *The Recognition of Declarative Questions in Information Dialogues*. PhD Dissertation, Tilburg University.

BEUN, R.J. (1991) A Framework for Cooperative Dialogues. In: Taylor, M.M., Néel, F. & Bouwhuis, D.G. (eds.) *Proceedings of the Second Venaco Workshop on the Structure of Multimodal Dialogue*. Maratea, Italy.

BEUN, R.J. (1994) Mental State Recognition and Communicative Effects. *Journal of Pragmatics, 21*. 191-214.

BEUN, R.J. & AHN, R.M.C. (1995) The DenK-system: modelling pragmatic issues in interaction with machines. *IPO Annual Progress Report 29*, 87-94.

VAN BENTHEM, J. (1991) Reflections on Epistemic Logic. *Logique et Analyse, 133-134*. 5-14.

BORGHUIS, T. (1993) Interpreting Modal Natural Deduction in type theory. In: de Rijke, M. (ed.) *Diamonds and Defaults*. Kluwer Academic Publishers. 67-102.

BORGHUIS, T. (1994) *Coming to Terms with Modal Logic: On the Interpretation of Modalities in Typed λ-Calculus*. Ph.D. Dissertation, Eindhoven University of Technology.

BORGHUIS, T. (1994) De integratie binnen het DenK-programma: overzichtsrapport project DK4. Samenwerkingsorgaan Brabantse Universiteiten, Tilburg/Eindhoven.

DE BRUIJN, N.G. (1980) A survey of the project Automath. In: Seldin & Hindley (eds.) *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalisms*. Academic Press. 579-606.

BUNT, H.C. (1989) Information dialogues as communicative action in relation to partner modelling and information processing. In: Taylor, M.M., Néel, F. & Bouwhuis, D.G. (eds.) *The Structure of Multimodel Dialogue*. North-Holland, Amsterdam. 47-73.

BUNT, H.C. (1991) DIT - Dynamic Interpretation in Text and Dialogue. In L. Kálm'an & L. Pólos (eds.) *Papers from the Second Symposium on Logic and Language*, Budapest: Akademiai Kiadó.

BUNT, H.C. (1993) Dynamic interpretation and dialogue theory. In: Black, W.J. & Bunt, H.C. (eds.) *Abduction, Belief and Context*, London University Press. (to appear)

BUNT, H.C. (1994) Context and Dialogue Control. *THINK Quarterly 3(1), 19-31* .

BUNT, H.C. (1995a) Dialogue Control Functions and Interaction Design. In Beun, R.J., Baker, M. and Reiner, M. (eds.) Dialogue and Instruction: Modelling Interaction in Intelligent Tutoring Systems. Berlin: Springer Verlag. (to appear)

BUNT, H.C. (1995b) Dialogue Semantics and Pragmatics in the ΔELTA Project. To appear in: N.O. Bernsen & L. Dybkjaer (eds.) *Papers from the ELSNET HCM Workshop on Dialogue and Discourse, Dublin, April 1995.*

BUNT, H.C. & BEUN, R.J. (1992) DenK: Dialogue Management and Knowledge Acquisition. *THINK 1(1).*

BURGER, J.D. & MARSHALL, R.J. (1993) The application of natural language models to intelligent multimedia. In: Mayburry, M. (ed.) *Intelligent Multimedia Interfaces.* Massachusetts: The MIT Press. 174-196.

CARPENTER, R. (1982) *The Logic of Typed Feature Structures.* Cambridge: Cambridge University Press.

COQUAND, T. (1985) *Une théorie des Constructions.* Thèse de troisième cycle. Paris: Université de Paris VII.

Referring in a shared workspace. In: Brouwer, M. & Harrington, T. (eds.) *Basics of Man-Machine Communication for the Design of Educational Systems.* Berlin: Springer-Verlag. 71-78.

CREMERS, A.H.M. (1995) Object reference during terminal dialogues. *This volume.*

CURRY, H.B. & FEYS R. (1958) *Combinatory Logic, Vol. 1,* North Holland Publishing Company.

GEURTS, B. & RENTIER, G. (1993) The PLUS Grammar. Internal Report ESPRIT Project PLUS, Institute for Language Technology and Artificial Intelligence ITK, Tilburg.

GRICE, H.P. (1975) Logic and Conversation. In: Cole, P. & Morgan, J. (eds.) *Speech Acts. Syntax and semantics, Vol. 11.* New York: Academic Press. 41-58.

HELMINK, L. & AHN, R.M.C (1991) Goal Oriented proof construction in type theory. In: Huet, G. and Plotkin, G. (eds.) *Logical Frameworks.* Cambridge: Cambrige University Press. 120-148.

HUTCHINS, E. (1989) Metaphors for Interface Design. In: Taylor, M.M., Néel, F. & Bouwhuis, D.G. (eds.) *The Structure of Multimodal Dialogue.* Amsterdam: North-Holland. 11-28.

JASPARS, J.O.M. (1993) *Calculi for Constructive Communciation.* Ph.D. Dissertation, Institute for Language Technology and Artificial Intelligence ITK, Tilburg.

KAMP, J.A.W. (1981) A Theory of Truth and Semantic Representation. In: Groenendijk, J. & Stokhof, M. (eds.) *Formal Methods in the Study of Language.* Amsterdam: Mathematisch Centrum. 277-322.

KAMP, J.A.W. & REYLE, U. (1993) From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Dordrecht: Kluwer Academic Publishers.

KIEVIT, L.A. (1994) Proto-ULF. *DenK Working Paper, October 1994,* ITK, Tilburg University.

MÄENPÄÄ, P. & RANTA, A. (1990) An implementation of Intuitionistic Categorial Grammar. In: Kálmán, L. & Pólos, L. (eds.) *Papers from the Second Symposium on Logic and Language*, Akadémia Kiadó, Budapest. 299-318.

MARTIN-LÖF, P. (1984) *Intuitionistic Type Theory* . Naples: Bibliopolis.

MEYER, B. (1988) *Object-oriented software construction.* Prentice Hall International.

NEAL, J.G. & SHAPIRO, S.C. (1991) Intelligent multimedia interface technology. In: Sullivan, J.W. & Taylor, S.W. (eds.) *Intelligent User Interfaces.* Reading, MA: Addison-Wesley. 11-43.

VAN OVERVELD, C.W.A.M. (1991) The Generalized Display Processor as an approach to real time interactive 3-D computer animation. *The Journal of Visualisation and Computer Animation, 2(1).* 16-21.

PEETERS, E.A.J. (1994) LOOKS*: Syntax and Semantics.* Computing Science Note. Eindhoven University of Technology.

POLLARD, C. & SAG, I. (1987) *Information-based approach to syntax and semantics.* CSLI, Stanford.

RENTIER, G. (1993) ULF. ΔELTA *Working Paper, May 1994*, ITK, Tilburg University.

SEARLE, J.R. (1969) *Speech acts.* Cambridge: Cambridge University Press.

DE SOUZA, C.S. (1993) The semiotic engineering of user interface languages. *International Journal of Man-Machine Studies, 39.* 753-774.

WAHLSTER, W., ANDRÉ, E., FINKLER, W., PROFITLICH, H.-J. & RIST, T. (1993) Plan-based integration of natural language and graphics generation. *Artificial Intelligence, 63.* 387-427.

WILSON, M.D., SEDLOCK, D., BINOT, J-L. & FALZON, P. (1991) An architecture for multimodal dialogue. In: Taylor, M.M., Néel, F. & Bouwhuis, D.G. (eds.) *Proceedings of the Second Venaco Workshop on the Structure of Multimodal Dialogue.* Maratea, Italy.

# Multimodal Maps: An Agent-based Approach

Adam Cheyer and Luc Julia

SRI International
333 Ravenswood Ave
Menlo Park, CA 94025 - USA

**Abstract**

In this paper, we discuss how multiple input modalities may be combined to produce more natural user interfaces. To illustrate this technique, we present a prototype map-based application for a travel planning domain. The application is distinguished by a synergistic combination of handwriting, gesture and speech modalities; access to existing data sources including the World Wide Web; and a mobile handheld interface. To implement the described application, a hierarchical distributed network of heterogeneous software agents was augmented by appropriate functionality for developing synergistic multimodal applications.

**Key words:** Multimodal Interface, Agent Architecture, Distributed Artificial Intelligence.

## 1 Introduction

As computer systems become more powerful and complex, efforts to make computer interfaces more simple and natural become increasingly important. Natural interfaces should be designed to facilitate communication in ways people are already accustomed to using. Such interfaces allow users to concentrate on the tasks they are trying to accomplish, not worry about what they must do to control the interface.

In this paper, we begin by discussing what input modalities humans are comfortable using when interacting with computers, and how these modalities should best be combined in order to produce natural interfaces. In section three, we present a prototype map-based application for the travel planning domain which uses a synergistic combination of several input modalities. Section four describes the agent-based approach we used to implement the application and the work on which it is based. In section five, we summarize our conclusions and future directions.

## 2 Natural Input

### 2.1 Input Modalities

Direct manipulation interface technologies are currently the most widely used techniques for creating user interfaces. Through the use of menus and a graphical user interface, users are presented with sets of discrete actions and the objects on which to perform them. Pointing devices such as a mouse facilitate selection of an object or action, and drag and drop techniques allow items to be moved or combined with other entities or actions.

With the addition of electronic pen devices, gestural drawings add a new dimension direct manipulation interfaces. Gestures allow users to communicate a surprisingly wide range of meaningful requests with a few simple strokes. Research has shown that multiple gestures can be combined to form dialog, with rules of temporal grouping overriding temporal sequencing [22]. Gestural commands are particularly applicable to graphical or editing type tasks.

Direct manipulation interactions possess many desirable qualities: communication is generally fast and concise; input techniques are easy to learn and remember; the user has a good idea about what can be accomplished, as the visual presentation of the available actions is generally easily accessible. However, direct manipulation suffers from limitations when trying to access or describe entities which are not or can not be visualized by the user.

Limitations of direct manipulation style interfaces can be addressed by another interface technology, that of natural language interfaces. Natural language interfaces excel in describing entities that are not currently displayed on the monitor, in specifying temporal relations between entities or actions, and in identifying members of sets. These strengths are exactly the weaknesses of direct manipulation interfaces, and concurrently, the weaknesses of natural language interfaces (ambiguity, conceptual coverage, etc.) can be overcome by the strengths of direct manipulation.

Natural language content can be entered through different input modalities, including typing, handwriting, and speech. It is important to note that, while the same textual content can be provided by the three modalities, each modality has widely varying properties.

- Spoken language is the modality used first and foremost in human-human interactive problem solving [4]. Speech is an extremely fast medium, several times faster than typing or handwriting. In addition, speech input contains content that is not present in other forms of natural language input, such as prosidy, tone and characteristics of the speaker (age, sex, accent).

- Typing is the most common way of entering information into a computer, because it is reasonably fast, very accurate, and requires no computational resources.

- Handwriting has been shown to be useful for certain types of tasks, such as performing numerical calculations and manipulating names which are difficult to pronounce [18, 19]. Because of its relatively slow production rate, handwriting may induce users to produce different types of input than is generated by spoken language; abbreviations, symbols and non-grammatical patterns may be expected to be more prevalent amid written input.

## 2.2 Combination of Modalities

As noted in the previous section, direct manipulation and natural language seem to be very complementary modalities. It is therefore not surprising that a number of multimodal systems combine the two.

Notable among such systems is the Cohen's Shoptalk system [6], a prototype manufacturing and decision-support system that aids in tasks such as quality assurance monitoring, and production scheduling. The natural language module of Shoptalk is based on the Chat-85

Figure 1: Multimodal Application for Travel Planning

natural language system [25] and is particularly good at handling time, tense, and temporal reasoning.

A number of systems have focused on combining the speed of speech with the reference provided by direct manipulation of a mouse pointer. Such systems include the XTRA system [1], CUBRICON [15], the PAC-Amodeus model [16], and TAPAGE [9].

XTRA and CUBRICON are both systems that combine complex spoken input with mouse clicks, using several knowledge sources for reference identification. CUBRICON's domain is a map-based task, making it similar to the application developed in this paper. However, the two are different in that CUBRICON can only use direct manipulation to indicate a specific item, whereas our system produces a richer mixing of modalities by adding both gestural and written language as input modalities.

The PAC-Amodeus systems such as VoicePaint and Notebook allow the user to synergistically combine vocal or mouse-click commands when interacting with notes or graphical objects. However, due to the selected domains, the natural language input is very simple, generally of the style "Insert a note here."

TAPAGE is another system that allows true synergistic combination of spoken input with direct manipulation. Like PAC-Amodeus, TAPAGE's domain provides only simple linguistic input. However, TAPAGE uses a pen-based interface instead of a mouse, allowing gestural commands. TAPAGE, selected as a building block for our map application, will be described more in detail in section 4.2.

Other interesting work regarding the simultaneous combination of handgestures and gaze can be found in [2, 13].

## 3 A Multimodal Map Application

In this section, we will describe a prototype map-based application for a travel planning domain. In order to provide the most natural user interface possible, the system permits the user to simultaneously combine direct manipulation, gestural drawings, handwritten, typed and spoken natural language When designing the system, other criteria were considered as well:

- The user interface must be light and fast enough to run on a handheld PDA while able to access applications and data that may require a more powerful machine.
- Existing commercial or research natural language and speech recognition systems should be used.
- Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web.

As illustrated in Figure 1, the user is presented with a pen sensitive map display on which drawn gestures and written natural language statements may be combined with spoken input. As opposed to a static paper map, the location, resolution, and content presented by the map change, according to the requests of the user. Objects of interest, such as restaurants, movie theaters, hotels, tourist sites, municipal buildings, etc. are displayed as icons. The user may ask the map to perform various actions. For example :

- *distance calculation* : e.g. "How far is the hotel from Fisherman's Wharf?"
- *object location* : e.g. "Where is the nearest post office?"
- *filtering* : e.g. "Display the French restaurants within 1 mile of this hotel."
- *information retrieval* : e.g. "Show me all available information about Alcatraz."

The application also makes use of multimodal (multimedia) output as well as input: video, text, sound and voice can all be combined when presenting an answer to a query.

During input, requests can be entered using gestures (see Figure 2 for sample gestures), handwriting, voice, or a combination of pen and voice. For instance, in order to calculate the distance between two points on the map, a command may be issued using the following:

- *gesture*, by simply drawing a line between the two points of interest.
- *voice*, by speaking "What is the distance from the post office to the hotel?".
- *handwriting*, by writing "dist p.o. to hotel?"
- *synergistic combination of pen and voice*, by speaking "What is the distance from here to this hotel?" while simultaneously indicating the specified locations by pointing or circling.

Notice that in our example of synergistic combination of pen and voice, the arguments to the verb "distance" can be specified before, at the same time, or shortly after the vocalization of the request to calculate the distance. If a user's request is ambiguous or underspecified, the system will wait several seconds and then issue a prompt requesting additional information.

The user interface runs on pen-equipped PC's or a Dauphin handheld PDA ([7]) using either a microphone or a telephone for voice input. The interface is connected either by

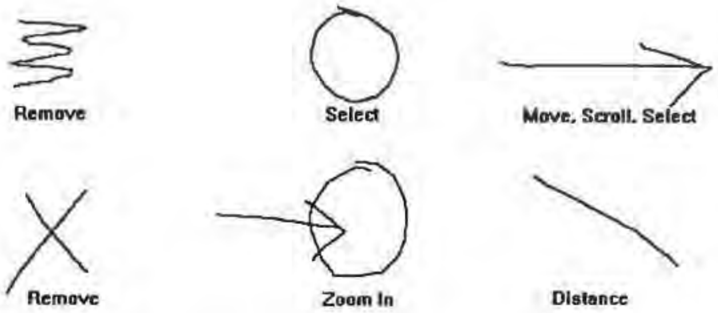


Figure 2: Sample gestures

modem or ethernet to a server machine which will manage database access, natural language processing and speech recognition for the application. The result is a mobile system that provides a synergistic pen/voice interface to remote databases.

In general, the speed of the system is quite acceptable. For gestural commands, which are handled locally on the user interface machine, a response is produced in less than one second. For handwritten commands, the time to recognize the handwriting, process the English query, access a database and begin to display the results on the user interface is less than three seconds (assuming an ethernet connection, and good network and database response). Solutions to verbal commands are displayed in three to five seconds after the end of speech has been detected; partial feedback indicating the current status of the speech recognition is provided earlier.

## 4 Approach

In order to implement the application described in the previous section, we chose to augment a proven agent- based architecture with functionalities developed for a synergistically multimodal application. The result is a flexible methodology for designing and implementing distributed multimodal applications.

### 4.1 Building Blocks

### 4.1.1 Open Agent Architecture

The Open Agent Architecture (OAA) [5] provides a framework for coordinating a society of agents which interact to solve problems for the user. Through the use of agents, the OAA provides distributed access to commercial applications, such as mail systems, calendar programs, databases, etc.

The Open Agent Architecture possesses several properties which make it a good candidate for our needs:

- An Interagent Communication Language (ICL) and Query Protocol have been developed, allowing agents to communicate among themselves. Agents can run on different platforms and be implemented in a variety of programming languages.
- Several natural language systems have been integrated into the OAA which convert English into the Interagent Communication Language. In addition, a speech recognition

agent has been developed to provide transparent access to the Corona speech recognition system.

- The agent architecture has been used to provide natural language and agent access to various heterogeneous data and knowledge sources.
- Agent interaction is very fine-grained. The architecture was designed so that a number of agents can work together, when appropriate in parallel, to produce fast responses to queries.

The architecture for the OAA, based loosely on Schwartz's FLiPSiDE system[23], uses a hierarchical configuration where client agents connect to a "facilitator" server. Facilitators provide content-based message routing, global data management, and process coordination for their set of connected agents. Facilitators can, in turn, be connected as clients of other facilitators. Each facilitator records the published functionality of their sub-agents, and when queries arrive in Interagent Communication Language form, they are responsible for breaking apart any complex queries and for distributing goals to the appropriate agents. An agent solving a goal may require supporting information and the agent architecture provides numerous means of requesting data from other agents or from the user.

Among the assortment of agent architectures, the Open Agent Architecture can be most closely compared to work by the ARPA knowledge sharing community [10]. The OAA's query protocol, Interagent Communication Language and Facilitator mechanisms have similar instantiations in the SHADE project, in the form of KQML, KIF and various independent capability matchmakers. Other agent architectures, such as General Magic's Telescript [11], MASCOS [20], or the CORBA distributed object approach [17] do not provide as fully developed mechanisms for interagent communication and delegation.

The Open Agent Architecture provides capability for accessing distributed knowledge sources through natural language and voice, but it is lacking integration with a synergistic multimodal interface.

### 4.1.2 TAPAGE

TAPAGE (edition de Tableaux par la Parole et la Geste) is a synergistic pen/voice system for designing and correcting tables.

To capture signals emitted during a user's interaction, TAPAGE integrates a set of modality agents, each responsible for a very specialized kind of signal [9]. The modality agents are connected to an "interpret agent" which is responsible for combining the inputs across all modalities to form a valid command for the application. The interpret agent receives filtered results from the modality agents, sorts the information into the correct fields, performs type-checking on the arguments, and prompts the user for any missing information, according to the model of the interaction. The interpret agent is also responsible for merging the data streams sent by the modality agents, and for resolving ambiguities among them, based on its knowledge of the application's internal state. Another function of the interpret agent is to produce reflexes: reflexes are actions output at the interface level without involving the functional core of the application.

The TAPAGE system can accept multimodal input, but it is not a distributed system; its functional core is fixed. In TAPAGE, the set of linguistic input is limited to a *verb object argument* format.

## 4.2 Synthesis

In the Open Agent Architecture, agents are distributed entities that can run on different machines, and communicate together to solve a task for the user. In TAPAGE, agents are used to provide streams of input to a central interpret process, responsible for merging incoming data. A generalization of these two types of agents could be :

*Macro Agents*: contain some knowledge and ability to reason about a domain, and can answer or make queries to other macro agents using the Interagent Communication Language.

*Micro Agents*: are responsible for handling a single input or output data stream, either filtering the signal to or from a hierarchically superior "interpret" agent.

The network architecture that we used was hierarchical at two resolutions – micro agents are connected to a superior macro agent, and macro agents are connected in turn to a facilitator agent. In both cases, a server is responsible for the supervision of its client sub-agents.

In order to describe our implementation, we will first give a description of each agent used in our application and then illustrate the flow of communication among agents produced by a user's request.

*Speech Recognition (SR) Agent*: The SR agent provides a mapping from the Interagent Communication Language to the API for the Decipher (Corona) speech recognition system [4], a continuous speech speaker independent recognizer based on Hidden Markov Model technology. This macro agent is also responsible for supervising a child micro agent whose task is to control the speech data stream. The SR agent can provide feedback to an interface agent about the current status and progress of the micro agent (e.g. "listening", "end of speech detected", etc.) This agent is written in C.

*Natural Language (NL) Parser Agent*: translates English expressions into the Interagent Communication Language (ICL). For a more complete description of the ICL, see [5]. The NL agent we selected for our application is the simplest of those integrated into the OAA. It is written in Prolog using Definite Clause Grammars, and supports a distributed vocabulary; each agent dynamically adds word definitions as it connects to the network. A current project is underway to integrate the Gemini natural language system [4], a robust bottom up parser and semantic interpreter specifically designed for use in Spoken Language Understanding projects.

*Database Agents*: Database agents can reside at local or remote locations and can be grouped hierarchically according to content. Micro agents can be connected to database agents to monitor relevant positions or events in real time. In our travel planning application, database agents provide maps for each city, as well as icons, vocabulary and information about available hotels, restaurants, movies, theaters, municipal buildings and tourist attractions. Three types of databases were used: Prolog databases, X.500 hierarchical databases, and data loaded automatically by scanning HTML pages from the World Wide Web (WWW). In one instance, a local newspaper provides weekly updates to its Mosaic-accessible list of current movie times and reviews, as well as adding several new restaurant reviews to a growing collection; this information is extracted by an HTML reading database agent and made accessible to the agent architecture. Descriptions and addresses of new restaurants are presented to the user on request, and the user can choose to add them to the permanent database by specifying positional coordinates on the map (eg. "add this new restaurant here"), information lacking in the WWW database.

*Reference Resolution Agent*: This agent is responsible for merging requests arriving in parallel from different modalities, and for controlling interactions between the user interface

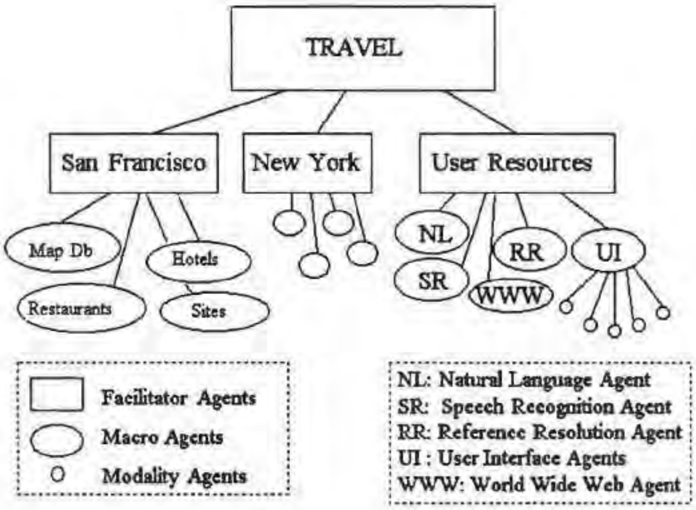


Figure 3: Agent Architecture for Map Application

agent, database agents and modality agents. In this implementation, the reference resolution agent is domain specific: knowledge is encoded as to what actions must be performed to resolve each possible type of ICL request in its particular domain. For a given ICL logical form, the agent can verify argument types, supply default values, and resolve argument references. Some argument references are descriptive ("How far is it to the hotel on Emerson Street?"); in this case, a domain agent will try to resolve the definite reference by sending database agent requests. Other references, particularly when contextual or deictic, are resolved by the user interface agent ("What are the rates for this hotel?"). Once arguments to a query have been resolved, this agent agent coordinates the actions and calculations necessary to produce the result of the request.

*Interface Agent:* This macro agent is responsible for managing what is currently being displayed to the user, and for accepting the user's multimodal input. The Interface Agent also coordinates client modality agents and resolves ambiguities among them : handwriting and gestures are interpreted locally by micro agents and combined with results from the speech recognition agent, running on a remote speech server. The handwriting micro-agent interfaces with the Microsoft PenWindows API and accesses a handwriting recognizer by CIC Corporation. The gesture micro- agent accesses recognition algorithms developed for TAPAGE.

An important task for the interface agent is to record which objects of each type are currently salient, in order to resolve contextual references such as "the hotel" or "where I was before." Deictic references are resolved by gestural or direct manipulation commands. If no such indication is currently specified, the user interface agent waits long enough to give the user an opportunity to supply the value, and then prompts the user for it.

We shall now give an example of the distributed interaction of agents for a specific query. In the following example, all communication among agents passes transparently through a facilitator agent in an undirected fashion; this process is left out of the description for brevity.

1. A user speaks: "How far is the restaurant from this hotel?"

2. The speech recognition agent monitors the status and results from its micro agent, sending feedback received by the user interface agent. When the string is recognized, a translation is requested.
3. The English request is received by the NL agent and translated into ICL form.
4. The reference resolution agent (RR) receives the ICL distance request containing one definite and one deictic reference and asks for resolution of these references.
5. The interface agent uses contextual structures to find what "the restaurant" refers to, and waits for the user to make a gesture indicating "the hotel", issuing prompts if necessary.
6. When the references have been resolved, the domain agent (RR) sends database requests asking for the coordinates of the items in question. It then calculates the distance according to the scale of the currently displayed map, and requests the user interface to produce output displaying the result of the calculation.

## 5 Conclusions

By augmenting an existing agent-based architecture with concepts necessary for synergistic multimodal input, we were able to rapidly develop a map-based application for a travel planning task. The resulting application has met our initial requirements: a mobile, synergistic pen/voice interface providing good natural language access to heterogeneous distributed knowledge sources. The approach used was general and should provide a for developing synergistic multimodal applications for other domains.

The system described here is one of the first that accepts commands made of synergistic combinations of spoken language, handwriting and gestural input. This fusion of modalities can produce more complex interactions than in many systems and the prototype application will serve as a testbed for acquiring a better understanding of multimodal input.

In the near future, we will continue to verify and extend our approach by building other multimodal applications. We are interested in generalizing the methodology even further; work has already begun on an agent-building tool which will simplify and automate many of the details of developing new agents and domains.

## References

[1] Allegayer, J, Jansen-Winkeln, R., Reddig, C. and Reithinger, N. "Bidirectional use of knowledge in the multi-modal NL access system XTRA". In Proceedings of IJCAI-89, Detroit, pp. 1492-1497.

[2] Bolt, R. "Put that there: Voice and Gesture at the Graphic Interface". Computer Graphics, 14(3), 1980, pp. 262-270.

[3] Bellik, Y. and Teil, D. "Les types de multimodalites", In Proc. IIM'92 (Paris), pp. 22-28.

[4] Cohen, M., Murveit, H., Bernstein, J., Price, P., Weintraub, M., "The DECIPHER Speech Recognition System". 1990 IEEE ICASSP, pp. 77-80.

[5] Cohen, P.R., Cheyer, A., Wang, M. and Baeg, S.C. "An Open Agent Architecture". In Proc. AAAI'94 - SA (Stanford), pp. 1-8.

[6] Cohen, P. "The role of natural language in a multimodal interface". Proceedings of UIST'92, 143-149.

[7] Dauphin DTR-1 User's Manual, Dauphin Technology, Inc. 337 E. Butterfield Rd., Suite 900, Lombard, Ill 60148.

[8] Dowding, J., Gawron, J.M., Appelt, D., Bear, J., Cherny, L., Moore, B. and Moran D., "Gemini: A natural language system for spoken-language understanding", Technical Note 527, AI Center, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025, April 1993.

[9] Faure, C. and Julia, L. "An Agent-Based Architecture for a Multimodal Interface". In Proc. AAAI'94 - IM4S (Stanford), pp. 82-86.

[10] Genesereth, M. and Singh, N.P. "A knowledge sharing approach to software interoperation". Computer Science Department, Stanford University, unpublished ms., 1994.

[11] General Magic, Inc., "Telescript Product Documentation", 1995.

[12] Julia, L. and Faure, C. "A Multimodal Interface for Incremental Graphic Document Design". HCI International '93, Orlando.

[13] Koons, D.B., Sparrell, C.J., and Thorisson, K.R. "Integrating Simultaneous Input from Speech, Gaze and Hand Gestures". In *Intelligent Multimedia Interfaces*, Edited by Mark Maybury, Menlo Park, CA, AAAI Press, 1993.

[14] Maybury, M.T. (ed.), *Intelligent Multimedia Interfaces*, AAAI Press/MIT Press: Menlo Park, Ca, 1993.

[15] Neal, J.G., and Shapiro, S.C. "Intelligent Multi-media Interface Technology". In *Intelligent User Interfaces*, Edited by J. Sullivan and S. Tyler, Addison-Wesley Pub. Co., Reading, MA, 1991.

[16] Nigay, L. and Coutaz, J. "A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion". In Proc. InterCHI'93 (Amsterdam), ACM Press, pp. 172-178.

[17] Object Management Group, "The Common Object Request Broker: Architecture and Specification", OMG Document Number 91.12.1, December 1991.

[18] Oviatt, S. "Toward Empirically-Based Design of Multimodal Dialogue Systems". In Proc. AAAI'94 - IM4S (Stanford), pp. 30-36.

[19] Oviatt, S. and Olsen, E. "Integration Themes in Multimodal Human-Computer Interaction". Proceedings of ICSLP'94, Yokohama, pp. 551-554.

[20] Park, S.K., Choi J.M., Myeong-Wuk J., Lee G.L., and Lim Y.H. "MASCOS : A Multi-Agent System as the Computer Secretary". Submitted for publication.

[21] Pfaff, G. and Ten Hagen, P.J.W. *Seeheim workshop on User Interface Management Systems* (Berlin), Springer- Verlag.

[22] Rhyne J. "Dialogue Management for Gestural Interfaces". Computer Graphics, 21(2), 1987, pp. 137-142.

[23] Schwartz, D.G. "Cooperating heterogeneous systems: A blackboard-based meta approach". Technical Report 93-112, Center for Automation and Intelligent Systems Research, Case Western Reserve University, Cleveland Ohio, April 1993. Unpublished PhD. thesis.

[24] Sullivan, J. and Tyler, S. (eds.), *Intelligent User Interfaces*, Addison-Wesley Pub. Co., Reading, MA, 1991.

[25] Warren, D. and Pereira, F., "An Efficient Easily Adaptable System for Interpreting Natural Language Queries", in American Journal of Computational Linguistics, 8(3), 1982, pp. 110-123.

[26] Wauchope, K., "Eucalyptus: Integrating Natural Language with a Graphical User Interface." Naval Research Laboratory Technical Report NRL/FR/5510-94-9711, in press, 1994.

# Object reference during task-related terminal dialogues

Anita Cremers

Institute for Perception Research
P.O. Box 513 5600 MB Eindhoven The Netherlands
e-mail: cremers@prl.philips.nl

### Abstract

In the DenK project a multimodal interface is being developed which is suitable for graphical interaction as well as communication by means of natural language. For the design of this interface knowledge is needed about how humans refer to objects in a task-related environment, by means of natural language as well as gestures. In this paper some results of an experiment on referring behaviour in task-related terminal dialogues are reported on, and compared to those of a preceding experiment on spoken dialogues. The differences that occurred between the two modalities were mainly related to the ease either to produce utterances, or to coordinate between using language, gesturing and inspecting the task domain or to change turns. These differences were all found to be based on the so-called principle of minimal cooperative total effort, i.e. within the limitations of the available modalities the participants tried to use as less effort as possible to, on the one hand, refer to a certain object, and, on the other hand, identify the object. On the basis of the results some recommendations are provided for the design of a multimodal interface including the possibility of interaction by means of typed natural language.

**Keywords:** object reference, gestures, minimal effort, focus of attention, multimodal interface, absolute features, relative features.

## 1 Introduction

In the so-called *DenK*-project[1] [Ahn et al., 1995], a multimodal interface is being developed which is suitable for graphical interaction as well as communication by means of natural language. The DenK interface can be represented as a triangle as shown in Figure 1. The angles of this triangle stand for the user, the domain and the *cooperative assistant*, of which the latter two are components of the interface. The domain can be seen as the collection of objects represented on the screen and the relations between them. The cooperative assistant can be seen as the user's collocutor who is also able to perform actions in the domain. The user is allowed to point at objects in the domain or manipulate them directly by means of some input device (e.g. a mouse). The user can also instruct the cooperative assistant by means of natural language to carry out certain actions in the domain, or ask questions about objects or events that play a role in the interaction.

---

[1] DenK stands for 'Dialoogvoering en Kennisopbouw' in Dutch, which means 'Dialogue Management and Knowledge Acquisition'. It is a joined research program of the universities of Tilburg and Eindhoven, and is partly financed by the Tilburg-Eindhoven Organisation for Inter-University Cooperation.

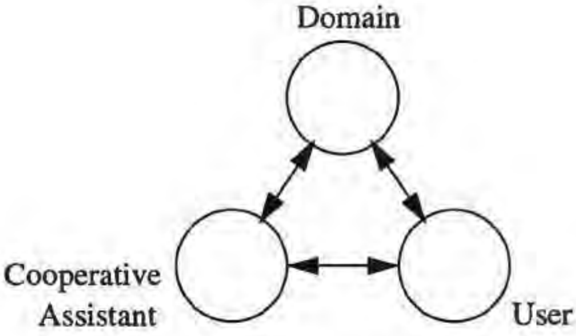


Figure 1: The DenK triangle

If a user wants to ask questions or give instructions, it is important to make clear which objects are involved. In a multimodal interface the act of referring to objects can be performed by means of either natural language or pointing or a combination of the two. In any case, the user should take care to provide appropriate information for the system to be able to identify the intended object (the *target object*).

To equip the system with knowledge of how humans refer to objects in a 'natural' situation, empirical research on this topic is needed. One of the most natural ways for humans to communicate is by means of speech. However, owing to technological limitations, most natural language systems only allow typed input. Unfortunately, it is not possible to extrapolate results from research on 'natural' spoken dialogues to written dialogues. It has been shown that there are notable differences between the two modes of communication, in particular with respect to length and syntax [Hauptmann & Rudnicky, 1988], the speed and the planning of utterances, and the nature of the speech acts used [Oviatt & Cohen, 1991]. For instance, more indirectness occurs in spoken dialogues than in terminal dialogues [Beun & Bunt, 1987]. In particular with respect to referential behaviour it was found, when referring to objects for the first time, that in telephone (spoken) dialogues more requests for identification occur than in keyboard dialogues [Cohen, 1984]. However, since this study dealt with telephone dialogues, only linguistic interaction was possible here. To conclude, to enable conclusions about referential behaviour in multimodal situations to be drawn, research on both spoken and typed dialogues is needed.

The referential behaviour of participants in spoken task-related dialogues in a situation designed to mimic the DenK triangle has already been investigated in a previous study [Cremers & Beun, 1995]. The present paper deals with an empirical study on how humans refer to objects in a similar type of *terminal* dialogue. The focus will lie on the type and amount of information humans use in referential expressions and the use of gestures. The results of this study will be compared with findings from the previous research on spoken dialogues based on the differences between the two situations as represented in the DenK triangle.

In section 2 some results from the previous study on spoken dialogues will be presented briefly. In section 3 some expectations will be formulated about findings in a corpus of terminal dialogues, based on the results obtained from the study on spoken dialogues and findings from the literature. In section 4 the results of checking the expectations in terminal dialogues will be presented and compared with the spoken dialogues. Finally, in section 5 the results will be discussed in the framework of DenK and some conclusions will be formulated.

## 2 Referential behaviour in spoken dialogues

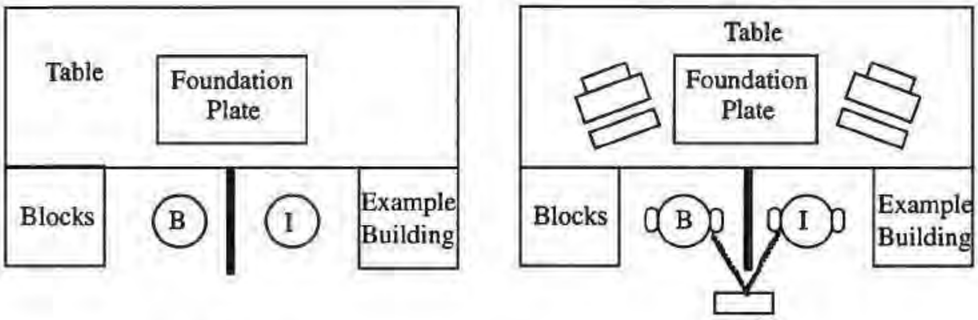
In a previous study on spoken dialogues [Cremers & Beun, 1995] an experiment was conducted to investigate the referential behaviour of ten pairs of participating subjects. The set-up of this experiment is depicted in Figure 2a. The study was designed to mimic the triangular DenK paradigm and can be described as follows. Two participants were seated side by side at a table but separated by a screen. To prevent communication other than by speech and gesturing, only the hands of each were visible to the other participant and then only when placed on top of the table. One of the participants (the instructor) was told to instruct the other (the builder) in reconstructing a block building on a toy foundation plate, placed on top of the table, in accordance with an example provided. In this set-up the role of the instructor was similar to that of the user and the role of the builder was similar to that of the cooperative assistant in the DenK triangle. Both participants were allowed to observe the building domain, to talk about it and to gesticulate in it, but only the builder was allowed to manipulate blocks. A brief overview of the main results of this experiment will be given in the subsections which follow.

### 2.1 The principle of minimal cooperative total effort

In the experiment on spoken dialogues participants were found to adhere to the so-called *principle of minimal cooperative total effort.* This principle expresses the idea that together the participants try to say [Clark & Wilkes-Gibbs, 1986] and do [Cremers & Beun, 1995] as little as possible, but just enough to be able to reach mutual agreement that the target object has been identified. For the speaker this means that he will transfer the least possible information and also a particular type of information to refer to the target object, so that it allows the hearer to identify the object by having to consider as few objects as possible. Consequences of this principle in the spoken dialogues were related to the *choice of features* in the referential expressions and the *focus of attention* of the participants.

The first consequence was that, if possible, speakers preferred to use *absolute features* rather than *relative features.* Absolute features such as the physical feature 'red' are features that can be understood by considering only the target object. Relative features can only be understood by also considering other objects or persons that are present. Relative features may be either implicit or explicit. To understand implicit relative features, such as the physical feature 'large', other objects have to be considered. To understand explicit relative features, such as the location feature 'to the right of', other objects have to be identified in order to permit identification of the target object. Absolute features are consequently easier to understand than relative features.

The second consequence was that speakers used less information to refer to objects located in the area of the building domain that was in the current focus of attention of the participants than to those located outside of this area. As part of the task changes had to be made in several parts of the block building. If changes are being made in a particular part of the building the speaker can assume that the focus of attention of both himself and his partner is directed at this area of the domain. For instance, participants used the referential expression 'the red block' to refer to the only red block within the current focus area, although many red blocks were present within the domain as a whole. Compared with the situation where the whole domain is taken into account, this means a reduction of words in the referential expression for the speaker, and fewer objects for the addressee to consider in order to find



(a) (b)

Figure 2: Experimental set-up for (a) spoken dialogues and (b) terminal dialogues.

the target object.

Furthermore, it was found that participants in choosing the next object preferred to refer to an object that was in the current focus of attention. This resulted in a larger proportion of references to objects in focus (68%) than to objects out of focus (32%). In terms of minimal effort this could be explained as a strategy to make optimal use of the current focus area before moving on to the next one.

### 2.2 The process of object reference

In the spoken dialogues there were usually some turn-takings before the participants arrived at the common agreement that the target object had been identified. It was found in [Cremers, 1994] that the number of turns needed was related to the focus of attention. To reach agreement on the identification of objects located within the current focus area fewer turns were needed than to refer to objects outside the current focus area (respectively 2.4 (s.d.=0.8) and 3.2 turns (s.d.=1.9)).

## 3 Terminal dialogues

In this section a description will first be given of the paradigm used for the study of multimodal terminal dialogues, followed by an overview of the differences between this paradigm and the previous one on spoken dialogues that was discussed in section 2. On the basis of the findings from the literature and from the preceding study on spoken dialogues some predictions for the outcome of this experiment will be formulated in section 3.2.

### 3.1 The experiment

A second experiment was carried out which was identical to that described in section 2, except for one important difference, namely that the participants communicated via keyboard and screen instead of by speech. In the DenK triangle this means that the mode of communication between the user and the cooperative asssistant is typed natural language. To prevent the participants from talking to each other instead of typing, they wore headphones to listen to some background music.[2] The experimental setup is depicted in Figure 2b.

---

[2] This set-up served its purpose since the 10 pairs of subjects who participated never spoke to each other at any time during the experiment.

The change from spoken communication to typed communication has some important expected consequences for the manner in which object reference can be carried out. First, the coordination between different modes of communication is expected to be different. In the spoken modality it is possible to speak and inspect the domain or point at objects in the domain at the same time. This is not possible in the terminal situation. If a participant is typing, his attention is directed at the screen and the keyboard, so that he cannot see what is going on in the block-building process. Also, since his hands are busy typing, he cannot use them to point at objects in the domain. A second consequence of the change from spoken to terminal dialogues is that it is more difficult to take turns. To pass the turn on to the partner a participant had to explicitly press a certain key. Only after he did so was the partner able to type. If a participant wanted to take his turn to type, he had to ask for it explicitly by means of a special key, and the partner had to acknowledge the switch of turn by pressing another key.[3]

Some expectations as to referential behaviour in terminal dialogues will now be formulated, based on the consequences of the use of typed communication instead of spoken communication.

## 3.2 Expectations for terminal dialogues

### 3.2.1 Expectations about minimal effort

A general prediction with respect to terminal dialogues that is an effect of the principle of minimal cooperative total effort is that it normally takes more effort to conduct a terminal dialogue than a spoken dialogue, due to characteristics of the communicative modalities that are available. This difference in effort will be reflected in the length of referential expressions, the features chosen in the referential expressions and the use of gestures.

It is known from the literature (e.g. [Oviatt & Cohen, 1991]) that written dialogues generally take longer and contain fewer words than spoken dialogues. These results were also expected in the present experiment. The latter expectation also follows from the principle of minimal cooperative total effort. Since it takes more effort to type than to speak, fewer words will be used when typing. Written dialogues take more time than spoken dialogues but this increase would probably be even larger if more words were typed. However, the increase in time is not due only to the increase in effort. It can also be a consequence of the fact that participants do not feel as pressed for time as in spoken dialogues, so they take more time to formulate their utterances [Beun & Bunt, 1987]. With respect to the use of referential expressions in terminal dialogues the participants are expected to try to utter the same information but use fewer words than in spoken dialogues. Probably also more gestures will be used, in order to compensate for the reduction in words.

With respect to the choice of features, the prediction is that, just as in spoken dialogues, participants will have a preference for using absolute features. There is no reason to assume that **more** absolute features will be used in terminal dialogues, since the process of understanding a referential expression and identifying the referent is the same in both situations. An effect is, however, expected in the coordination of language and gestures. Since it is not possible to type and gesture at the same time, pointing gestures accompanied by

---

[3] If the participants had been allowed to type at the same time, this would have caused problems for them, especially since actions in the domain had to be monitored as well. In particular, the order of the turns and actions would have been less obvious.

demonstratives are expected to occur less often in terminal dialogues.

As a result of a general reduction in words and an expected increase in the use of gestures, some of the features that were used in spoken dialogues will have to be replaced by gestures in typed dialogues. Tentative predictions are that absolute features containing information that cannot be expressed very easily by gestures (e.g. colour) will continue to be used, but that the rather verbose explicit relative features will be replaced by gestures.

The reduction of words as a result of the current focus of attention is expected to occur more often in terminal dialogues than in spoken dialogues. A reduction of words means less typing and therefore less effort on the part of the participant. However, since the coordination of typing and inspecting the domain at the same time is difficult in terminal dialogues, it is expected that participants will easily loose track of the current focus area. This will probably result in a relatively smaller number of references to objects in focus than in the spoken dialogues.

### 3.2.2 Expectations about the process of object reference

In the spoken dialogues it was very easy to react immediately to something the partner said, resulting in a mean number of turns of 2.7 before mutual agreement was reached that the target object had been identified. The prediction for terminal dialogues is that the effort to take turn to type will be so large that in most of the dialogues hardly any verbal turn-takings will take place. First, this could mean that more information will be given in the first turn, to avoid having to use more verbal turns. Note that this expectation contradicts the expected general reduction of words in referential expressions in terminal dialogues. A second possible consequence is that the reduction in verbal turns will be compensated for by an increase in non-verbal turns since there is no inherent difficulty in taking turns in gesturing during terminal dialogues.

There could be a reason for a possible *in*crease in verbal turns as well. This increase could be a result of the occurrence of more *miscommunications* during the terminal dialogues, although it is suggested in the literature (see [Cohen, 1984]) that this effect does not exist. A miscommunication is defined as an event whereby a wrong selection takes place before the right target object is identified. The expectation of an increase in miscommunications is a consequence of the expected decrease in words in terminal dialogues. To correct the miscommunication and identify the right target object additional turns will be needed. However, if the expectation about giving more information in the first turn to avoid having to engage in tedious turn-takings is correct, an increase in miscommunications is not likely to occur.

Finally, it is not clear whether in the terminal dialogues, as in the spoken dialogues, the number of turns to refer to objects in focus will be lower than those to refer to objects out of focus. In terminal dialogues, where participants have to divide their attention between keyboard, screen and domain, it is harder for them to continue focusing their attention on the current focus area. This could mean that they will not succeed in benefiting from the focus area as much as the participants in spoken dialogues did. In other words, it is probable that no difference in the number of turns will occur between in focus and out of focus.

## 4 Results

In the terminal dialogues a total number of 156 referential acts occurred, which is almost the same as the number of referential acts found in the spoken dialogues, namely 145. This result

| | *terminal* | *spoken* |
|---|---|---|
| *mean length* | 12 min.<br>(189 words) | 4 min. 47 sec.<br>(729 words) |
| *mean length of language* | 7 min. 56 sec.<br>(0.4 words/sec.) | 4 min. 17 sec.<br>(2.8 words per sec.) |
| *mean length of actions* | 4 min. 4 sec.<br>(34% of total time) | 30 sec.<br>(10% of total time) |

Table 1: Mean length of terminal and spoken dialogues

is not surprising since both experiments involved exactly the same task and the same objects.

Findings with respect to the principle of minimal cooperative total effort and the process to reach mutual agreement on identification will now be discussed, and compared with the spoken dialogues.

## 4.1 Results concerning minimal effort

### 4.1.1 Length

**Length of the dialogues** In the literature it has been stated that, generally speaking, fewer words are used and more time is needed in terminal dialogues than in spoken dialogues ([Oviatt & Cohen, 1991], [Beun & Bunt, 1987]). This was also found in the present study (see Table 1). The participants took a mean time of 12 minutes to complete the terminal dialogues, during which time they used 189 words. It took the participants a mean time of only 4 minutes and 47 seconds to complete the spoken dialogues, but in that time they used 729 words.

However, not all of the time was devoted to typing or speaking. A part of the time was used to carry out actions as well. The actions carried out were both pointing actions and manipulations within the domain. In the terminal dialogues, 7 minutes and 56 seconds were taken for the actual typing, which means that the typing rate was 0.4 words per second. In the spoken dialogues, 4 minutes and 17 seconds were used for speaking, which yields a speaking rate of 2.8 words per second.

The figures show that in terminal dialogues a relatively large part of the time was devoted to actions only, namely 4 minutes and 4 seconds, which is 34% of the time. In spoken dialogues 30 seconds were used for performing actions only, and that is 10% of the total time.

The above results show that, indeed, it takes more time to conduct a terminal dialogue than a spoken dialogue, under exactly the same conditions. In fact, it takes exactly seven times longer to type a word than to utter it. Also, the amount of time spent on carrying out actions is different for the two types of dialogue. In terminal dialogues over three times longer is spent carrying out actions than in spoken dialogues. Since the task in the two experiments was exactly the same, this result cannot be explained by a difference in manipulating objects in the domain. The dissimilarity is therefore probably due to an increase in the use of referential actions, i.e. pointing or other gestures to indicate an object in the domain.

| *length* | *terminal* | *spoken* |
|---|---|---|
| *0* | 46% | 15% |
| *1* | 12% | 46% |
| *2* | 15% | 16% |
| *3* | 9% | 2% |
| *4* | 3% | 6% |
| *5* | 3% | 6% |
| *more* | 12% | 9% |

Table 2: Number of content words in referential acts

**Length of the referential expressions** A more specific hypothesis is concerned with the length of referential acts used in terminal and spoken dialogues. The prediction was that, since fewer words are used in terminal dialogues than in spoken dialogues, the length of referential acts in terminal dialogues would also be shorter. This prediction did not completely prove true. Although the mean number of content words (i.e. all words except the determiner) used in terminal dialogues was 1.8 (s.d. = 2.53) , compared to 2.2 (s.d. = 2.69) in spoken dialogues, this difference does not mean that most references in terminal dialogues were shorter than in spoken dialogues. First, the standard deviations are too large to show a clear difference in length between the two types of dialogue. Second, similar percentages of all lengths of referential expressions occurred in both dialogues, except for the referential expressions of lengths 0 and 1 (see Table 2). More content-less referential acts, i.e. gestures or demonstratives or combinations of

these, occurred in terminal than in spoken dialogues (terminal: 46%, spoken: 15%). In contrast, fewer referential acts containing only one content word occurred (terminal: 12%, spoken: 46%).

These figures seem to indicate that at times when typists use gestures only, or gestures accompanied by a demonstrative expression, speakers use one feature, possibly accompanied by a gesture, and vice versa. Since no large reduction of words in referential expressions could be demonstrated, the total reduction of words in terminal dialogues must be due to a reduction of words in the remaining part of the utterances, i.e. the part where the action to be carried out is expressed.

However, if we do not count the number of words in the referential expressions but the referential expressions in which features are used a clear difference can be found. In terminal dialogues fewer features (either absolute or relative or both) were used than in spoken dialogues, namely in 56% and 85%, respectively, of the referential expressions (see Table 3). This result is mainly due to the fact that in terminal dialogues far more gestures without any language were used than in spoken dialogues, namely in 44% and 4%, respectively, of the references. Contrary to expectations, no difference could be found with respect to the total number of gestures used in terminal and spoken dialogues. In both types of dialogue the percentage was exactly the same, viz. 53%.

### 4.1.2 Features and gestures

| | *terminal (156)* | | *spoken (145)* | |
|---|---|---|---|---|
| | *+gesture* | *–gesture* | *+gesture* | *–gesture* |
| *absolute* | 9 (6%) | 38 (24%) | 45 (31%) | 43 (30%) |
| *relative* | – – | 2 (1%) | 2 (1%) | 2 (1%) |
| *abs. & rel.* | 4 (3%) | 35 (22%) | 7 (5%) | 24 (17%) |
| *demonstr.* | – – | – – | 17 (12%) | – – |
| *gesture only* | 68 (44%) | – – | 5 (4%) | – – |
| *Total* | 81 (53%) | 75 (47%) | 76 (53%) | 69 (47%) |

Table 3: Features and gestures used in terminal and spoken dialogues

**Preference for absolute features** One of the findings relating to the principle of minimal cooperative total effort in terminal dialogues is, not surprisingly, that participants do have a preference for using absolute features rather than relative features, as is shown in Table 3. Absolute features only were used in 47 cases (30%). In spoken dialogues absolute features only were used in 88 (61%) of the referential acts. The use of relative features was more or less the same in both types of dialogue, viz. two (1%) in terminal dialogues and four (2%) in spoken dialogues. Also, combinations of absolute and relative features occurred equally often in terminal and spoken dialogues, viz. 39 (25%) and 31 (22%), respectively.

At first sight it may seem surprising that fewer absolute features were used in terminal dialogues than in spoken dialogues. This seems to weaken the principle of minimal cooperative total effort. The solution to this problem lies in the use of gestures. If we assume that the use of gestures only or gestures combined with demonstratives is a means to use less effort, then the figures for the choice of features in terminal and spoken dialogues become very similar. For terminal dialogues this would mean that the referential acts which involve the least effort are those in which gestures only are used plus those in which only absolute feature are used. These two percentages add up to 74%. In spoken dialogues, summation of the numbers of referential acts by means of gestures only, gestures plus demonstratives and absolute features only amounts to 77%.

To summarize, participants both in terminal and in spoken dialogues try to reduce effort by choosing particular features. However, the choice of features is different in both types of dialogue. In terminal dialogues relatively more gestures only are used and in spoken dialogues relatively more absolute features only.

**Coordination of typing and gesturing** The expectation with respect to the coordination of typing and gesturing was that in terminal dialogues fewer demonstratives accompanied by gestures would occur. This indeed turned out to be true. In terminal dialogues no cases at all occurred, whereas in spoken dialogues this combination occurred in 17% of the cases. This difference could even be extended to the use of absolute features accompanied by gestures. In terminal dialogues they were used in 6% of cases, whereas in spoken dialogues they occurred in 31% of cases. Relative features and combinations of absolute and relative features accompanied by gestures occurred equally often in terminal as in spoken dialogues.

**Type of features and gestures** The prediction concerning continuation of the use of features that cannot be expressed by means of gestures proved correct. In both types of dialogues almost the same percentage of absolute colour features was used (terminal: 100%, spoken: 97% of the absolute features used). However, there was a difference in the use of absolute shape features (e.g. 'square'). In terminal dialogues 46% of the absolute features contained shape information, whereas in spoken dialogues this was the case in only 17%. A possible explanation for this difference is the fact that in spoken dialogues absolute features were about 4 times more often accompanied by gestures than in terminal dialogues (terminal: +gesture 9%, –gesture 46%; spoken: +gesture 36%, –gesture 47%). Since the use of pointing gestures makes the use of shape information superfluous, this type of information is probably used less in terminal dialogues. The feature 'colour' is probably so salient that participants tend to keep on using it, even though the

use of a pointing gesture makes it superfluous.

The use of relative features in both types of dialogues was almost the same (terminal: 1%, spoken: 2%). Although the number explicit relative features in terminal dialogues was lower than in spoken dialogues (terminal: 23%, spoken: 39%) no clear difference was found. However, there was a difference in relative features that were used to refer to locations within the domain. If a location in the domain is indicated this generally takes relatively more words than if only physical features of objects are mentioned. It could be shown that in spoken dialogues more relative features were used to refer to locations (91% of the relative features used) than in terminal dialogues (68%). This suggests that participants in terminal dialogues tend to avoid these relatively long expressions, and probably point instead.

### 4.1.3 Focus of attention

In the terminal dialogues 86 out of 156 referential acts were used to refer to objects in the current focus of attention (55%). The 70 remaining referential acts (45%) were used to refer to objects outside of the current focus area (see [Cremers & Beun, 1995] for the criteria used to make this bipartition). Hence, no clear preference for choosing the next object in or out of the current focus area could be detected, as was the case in the spoken dialogues (68% in focus, 32% out of focus). This result confirms the expectation and is probably due to a coordination problem between typing and inspecting the domain.

Among the 86 references used in the terminal dialogues to refer to objects within the current focus of attention, focus reduction was applied in 20 cases (23% of 86). This percentage is very close to that found in spoken dialogues, where focus reduction was applied in 27% of the cases. Our prediction was, however, that in terminal dialogues more cases of focus reduction would occur owing to a general reduction of words. The result seems to suggest that this was not the case. However, if we again consider the use of gestures as a means to reduce effort, some evidence for the truth of the hypothesis can be found.

Participants in terminal dialogues used gestures without any language to refer to objects in 35 (41%) of the in-focus cases. In spoken dialogues this was done in 13 cases (13%), where the gesture was accompanied by just a demonstrative. If we add the cases of gesture-related focus reduction to those where only a verbal reduction took place, the total number of cases of focus reduction in terminal dialogues becomes 55 (64% of the in-focus cases). In spoken dialogues the total number of focus reduction then becomes 40 (40% of the in-focus cases). This suggests that, in the latter interpretation of focus reduction, participants in terminal dialogues indeed use more reduced information when referring to objects within the focus area than do participants in spoken dialogues. However, this reduction is due more to the the use of gestures than to the use of reduced verbal information. An overview of the findings is

| | *terminal*<br>(86) | | *spoken*<br>(99) | |
|---|---|---|---|---|
| *verbal* | 23% | (20) | 27% | (27) |
| *gestures* | 41% | (35) | 13% | (13) |
| *Total* | 64% | (55) | 40% | (40) |

Table 4: Focus reduction in terminal and spoken dialogues

given in Table 4.

## 4.2 Results concerning the process of object reference

### 4.2.1 Number of turns

In terminal as well as spoken dialogues the mean number of both verbal and non-verbal turns needed to arrive at the mutual agreement that the target object has been identified is exactly the same, namely 2.7 (s.d. 1.04 and 1.38, repectively). However, this does not mean that the process is exactly the same for both types of dialogues. The difference lies in the relative use of verbal turns and (referential) actions in this process. In terminal dialogues 98 (63%) of the turns were non-verbal, whereas in spoken dialogues gestures or actions were used only in 23 (16%) of the turns. No indication was found that more information was given in the first turn to avoid turn-takings since the mean lengths of first referential acts in terminal and spoken dialogues were very similar (terminal: 1.8, spoken: 2.2) and even shorter in terminal dialogues.

With respect to the number of turns necessary to refer to objects in or out of focus a difference between spoken dialogues and terminal dialogues was found. In spoken dialogues more turns were needed to refer to an object out of focus (3.2) than to one in focus (2.4), whereas no difference could be found in terminal dialogues (both 2.7). This confirms our expectation that participants in terminal dialogues do not benefit very much from the focus area, probably due to coordination problems between typing and inspecting the domain.

### 4.2.2 Miscommunications

One of the expectations presented in section 3.2.2 was that in terminal dialogues more turns due to miscommunications would occur, since participants use fewer words to refer to objects. In the preceding section it was shown that no difference in the mean number of turns between terminal dialogues and spoken dialogues occurred. This means that, if more miscommunications occurred, they did not increase the mean number of turns significantly. The results of analysing the occurring miscommunications is given in Table 5.

In terminal dialogues miscommunications occurred in 25 (16%) of the cases before identification took place. In spoken dialogues only six (4%) of the first references to objects in the domain were initially misunderstood. These miscommunications were found to be due mainly to misunderstandings related to focus (in five cases, 83%). The one remaining case (17%) was due to a mistake made by the speaker.

In the terminal dialogues 13 (65%) of the misunderstandings were in some way related to focus. In four cases (16%) mistakes were made by either one of the participants. In the

| | *terminal* | *spoken* |
|---|---|---|
| *Total* | 25 (16%) | 6 (4%) |
| *focus* | 13 (65%) | 5 (83%) |
| *mistake* | 4 (16%) | 1 (17%) |
| *determiner* | 8 (19%) | – |
| *focus-det.* | 13 (77%) | 5 (83%) |

Table 5: Miscommunications in terminal and spoken dialogues

remaining eight cases (19%) the misunderstanding was a result of confusion as to whether a new object should be introduced or the referential act was meant to refer to an object in the domain. These confusions were directly related to the fact that the typists did not add any determiner to the referential expression. This is a clear consequence of the modality of communication that was used. In order to type as few words as possible, typists omitted determiners thereby leading to a misunderstanding.

Since the latter group of misunderstandings was a direct result of the available modalities of communication, they can be omitted from the comparison between terminal and spoken dialogues. The percentage of misunderstandings due to focus then becomes 77% (13 out of 17 cases), which is close to the 83% found in spoken dialogues.

To summarize, more or less the same percentage of focus-related misunderstandings occurred in terminal dialogues as in spoken dialogues. However, the total percentage of misunderstandings in terminal dialogues was greater since more misunderstandings occurred due to mistakes and, most importantly, due to omitting the determiner in the description. This result stresses the importance of determiners that provide information about the accessibility of the referent (see [Piwek & Cremers, 1995]).

## 5 Discussion and conclusions

The differences between the uses of referential expressions and gestures in terminal and spoken dialogues can be explained to a large extent by the differences in the respective experimental paradigms as illustrated by the DenK triangle.

A direct consequence of the change from spoken to typed communication is the length of the referential expressions used. Since it takes more effort to type than to speak, fewer words were used in referential expressions in terminal dialogues than in spoken dialogues. However, since the difference was not very great, the largest reduction of words occurred in the non-referential parts of the utterances. Furthermore, it could not be demonstrated that participants in terminal dialogues used fewer gestures than those in spoken dialogues. The total number of gestures was the same although the distribution over accompanying features was different. However, these results may be domain-dependent since objects that are more difficult to describe are expected to be pointed at more often.

The difference in the distribution of gestures was a direct consequence of the problematic coordination of verbal and non-verbal information in terminal dialogues. Since it was not possible to gesture and type at the same time, hardly any occurrences of short referential expressions, such as demonstratives or absolute features only, were found. In spoken dialogues the demonstratives and absolute features that accompanied gestures can be said to have the function of either attracting the attention of the partner to look at the domain or keeping the conversation flowing by avoiding silences. In terminal dialogues the latter function is not very prevalent since the time pressure is not so great there (see [Beun & Bunt, 1987]). Participants in terminal dialogues lost the possibility to apply the former function, i.e. to attract attention. However, these participants were observed to point with more emphasis, i.e. repeatedly or for a longer period than participants in spoken dialogues did. This emphasis can be interpret

ed as a means to make sure that the partner has observed the gesture.

A second consequence related to the coordination of modalities was the fact that typing and simultaneously inspecting the domain was difficult. This resulted in difficulty in keeping track of the current focus area. This difficulty was reflected in the same number of references to objects in focus to objects out of focus, compared to this distribution in spoken dialogues where far more references to objects in focus occurred.

As a consequence of the difficulty in changing turns in terminal dialogues fewer verbal turns took place. However, the loss of verbal turns was compensated by more non-verbal turns. There was no indication that more information was given in the first utterance to try to avoid having to use more turns. However, this could be a consequence of the relatively simple objects used in the experiment. It was probably not necessary to use more words to indicate a certain object unambiguously. Although more miscommunications occurred in an absolute sense, they did not affect the mean number of turns used to reach common agreement that the target object had been identified.

The differences between terminal and spoken dialogues were all found to be based on the principle of minimal cooperative total effort. In a situation where different modalities of communication are available which have different characteristics and possibilities, other means have to be found to minimize effort. The main change with respect to spoken dialogues was in the use of gestures to refer to objects. In both spoken dialogues and terminal dialogues the same numbers of gestures were used, although they were used at different moments. At moments where participants in spoken dialogues used limited information, participants in terminal dialogues tended to use more pointing gestures.

¿From these findings some implications can be drawn for the design of a multimodal interface, such as the DenK interface. First, in our domain we did not find a large reduction of words in referential expressions, but we did find a large reduction in the rest of the utterances, i.e. in the part were the action that has to be carried out is formulated. Further research should be conducted to figure out whether this reduction causes more or other types of miscommunications.

In the design of a multimodal interface special attention should be devoted to the coordination of verbal and non-verbal information. Procedures should be developed to make links between verbal expressions, especially longer ones, and gestures that are meant to refer to the same objects but do not occur at the same time. This is necessary in order to avoid confusions about whether in these cases only one object or two separate ones are being referred to.

In terminal dialogues participants apparently did not make use of the current focus area as often as participants in the spoken dialogues, but reduced expressions referring to objects in the current focus area still occurred regularly. This means that the interface should adopt a notion of focus area in order to enable these expressions to be understood.

Finally, the interface should allow users to change turns quickly since almost the only type of feedback that was provided in the terminal dialogues consisted of gestures or actions in the domain. It is probably easier for the interface to understand verbal feedback than to have to analyse the meaning of the gestures and actions. However, provisions should be made for listing the verbal and non-verbal turns in a convenient way so that no confusions will arise because the correct order of the turns is unclear.

## References

[Ahn et al., 1995] R.M.C. Ahn, R.J. Beun, T. Borghuis, H.C. Bunt & C.W.A.M. van Overveld. The DenK-architecture: a fundamental approach to user-interfaces. *Artificial Intelligence Review.*

[Beun & Bunt, 1987] R.J. Beun & H.C. Bunt. Investigating linguistic behaviour in information dialogues with a computer. *IPO Annual Progress Report 22.*

[Clark & Wilkes-Gibbs, 1986] H.H. Clark & D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition* 22, 1-39, 1986.

[Cohen, 1984] P.R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics 10(2)*, 97-125.

[Cremers, 1994] A.H.M. Cremers. The process of cooperative object reference in a shared domain. Submitted for *Proceedings of the workshop on form and function in grammatical theory.* First International Summer Institute in Cognitive Science, State University of New York at Buffalo (USA), 5-30 July 1994.

[Cremers & Beun, 1995] A.H.M. Cremers & R.J. Beun. *Object reference in a shared domain of conversation.* Eindhoven, Institute for Perception Research, manuscript no. 1089.

[Hauptmann & Rudnicky, 1988] A.G. Hauptmann & A.I. Rudnicky. Talking to computers: an empirical investigation. *International Journal of Man-Machine Studies 28*, 583-604, 1988.

[Oviatt & Cohen, 1991] S.L. Oviatt & P.R. Cohen. Discourse structure and performance efficiency in interactive and non-interactive spoken modalities. *Computer Speech and Language 5*, 297-326, 1991.

[Piwek & Cremers, 1995] P. Piwek & A. Cremers. Demonstratives in Dutch cooperative task dialogues: An analysis of the use and function of Dutch and English distal and proximate demonstratives. Eindhoven, Institute for Perception Research, *IPO report no. 1036.*

# Speakers' Responses to Requests for Repetition in a Multimedia Language Processing Environment

Laurel Fais, Kyung-ho Loken-Kim and Young-Duk Park

ATR Interpreting Telecommunications Res. Labs.
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02 Japan
email: fais@itl.atr.co.jp

## Abstract

This paper investigates the linguistic and modal aspects of responses made by subjects in a Wizard of Oz experiment to clarification requests made by the "Wizard." English-speaking "clients" participating in a task-oriented cooperative dialogue with Japanese-speaking "agents" were asked to clarify utterances that were complex or lengthy. The discourse, syntactic, and modal structures of these clarifications are examined. While linguistic factors are characterizable as "reducing" and "converging," media use in these responses does not exhibit a clear pattern. Implications are drawn for future investigations into the use of multimedia configurations and for the integration of multimedia technologies in automatic speech processing.

**Key words**: multimedia communication, automatic machine translation, Wizard of Oz experiment, bilingual cooperative dialogue.

## 1 Overview

Natural language processing systems are beginning to approach the difficult goal of handling unconstrained spontaneous speech. One way to improve the performance of such systems in this context is to supplement their processing capabilities with multimedia technologies designed to lessen the burden on the processing system. But the optimal configuration of supplemental media is not yet well understood; even less clear is the nature of the speech behavior we can expect from humans using multimedia speech processing systems. One thing is clear, however, from earlier work [1]: completely unconstrained spontaneous speech is likely to be too difficult to process entirely automatically for some time. Systems will have to be able to request, and receive, clarifications of users' utterances [2,3].

This work reports on a Wizard of Oz experiment in which English-speaking and Japanese-speaking subjects took part in a cooperative, task-oriented dialogue via a supposed "automatic machine translation system," (i.e., the Wizard), in two communication conditions: telephone-only and multimodal (MM), using the Environment for MultiModal Interaction (EMMI) designed and built at the Advanced Telecommunications Research Institute in Kyoto, Japan (ATR) [4]. It examines English-speaking subjects' clarification utterances and describes the ways in which these speakers accommodated, both linguistically and modally, to breakdowns in the "machine's" understanding.

There are a number of reasons why the nature of this behavior should be of interest. Most superficially, the fact that automatic processing systems dealing with spontaneous speech currently perform less than perfectly implies that requests by the machine for clarification from the human are a necessary feature of such systems. Knowledge of what strategies speakers are apt to employ in their clarification utterances can be used to enhance the ability of the processing system both to interpret the clarification itself and to situate it in the ongoing discourse. If we can specify a consistent and predictable relationship between the discourse, syntactic and lexical structures of pre- and post-clarification request utterances, we can use even partial information from the processing of the initial utterance to process the clarification utterance, increasing our chances of characterizing the speaker's contribution correctly.

Further, an examination of pre- and post-repetition request utterances reveals what modifications speakers can be expected to make *voluntarily* to instances of communication breakdown. We conjectured that speakers might slow their speaking rate, use fewer words, repeat a high percentage of words, or speak more fluently after a request for repetition. If this is the case, then, constraints that must be built into the communication environment so that the system can handle spontaneous speech more effectively can exploit these types of modifications either explicitly, through instruction, or implicitly, through discourse context [1]. It seems likely that encouraging those strategies that come naturally to speakers will be a more effective way to modify their communication behavior than trying to exploit strategies which are unfamiliar or, worse, difficult to carry out.

The investigation of these strategies in the context of a multimedia communication environment adds one further dimension to these issues. Does the availability and use of non-speech media have an effect on subjects' use of strategies that can lead to more easily processed spontaneous speech?

Finally, the nature of these accommodations is indicative of the aspects of speech behavior that speakers themselves feel generate "standard," understandable language. Studying speakers' responses to requests for clarification is similar to gaining an understanding of language by studying language pathologies such as stuttering or aphasia, or by examining disfluencies such as false starts or repairs.

While the latter is an interesting line to pursue, here we will focus on the more practical issues outlined above. That is, what is the relationship between the linguistic and modal behavior of speakers before and after repetition requests? What strategies do speakers use to modify their input in cases of communication breakdown? And finally, what are the implications of these results for automatic processing of spontaneous speech?

## 2 Methods

Twenty subjects, ten native speakers of American English and ten native speakers of Japanese, took part in the experiment. The English-speaking subjects, acting as "clients," were instructed that their task was to get directions to a specific place (the site of a conference they were supposedly attending) and to make a hotel reservation, by engaging in a cooperative dialogue with the Japanese-speaking "conference agents." All subjects were told that their speech would be translated by "ASyST," supposedly an "Automatic System for Speech Translation" which had been developed at ATR.

The "Wizards" for the experiment were experienced interpreters; a native American English speaker translated from Japanese to English and a native Japanese speaker translated from English to Japanese. These "Wizards" modulated their speech to be as monotonic and syllable-timed as possible, simulating the layman's impression of computer speech. The speech of both interpreters was passed through a Technics Mic Mixing Amplifier SH-3026 in order to make it sound more "machine-like" to the subjects. Each person taking part in the experiment, i.e., the two interpreters, the "client," and the "agent," could hear all of the speech produced by every other person. No subject indicated any doubt that his or her speech was being translated by a machine.

None of the subjects knew one another, nor were they at all familiar with EMMI. The subjects were told that they were to enact the experiment scenario twice, once via telephone and once via the multimedia interface. Five agent-client pairs participated in the telephone condition first; five used the multimedia set-up first.

In the MM condition, subjects sat in front of a NeXT computer monitor, with touchscreen, keyboard, and mouse. On the screen appeared a video image of the person with whom they were talking, a field for typing in written input, and an area in which several different maps or the hotel reservation form could be displayed by the agent. Subjects could draw on the map by dragging with the mouse or by hand, could type on the keyboard (activating the field by mouse or hand), or could use speech to communicate. Subjects were encouraged to practice with the drawing and typing capabilities of EMMI until they felt comfortable, and those acting as agent were thoroughly instructed in the information they had available to impart to the client.

In the telephone condition, subjects spoke into standard telephones. In both conditions, subjects wore Sennheiser HMD 410 headsets with microphone (one ear piece was turned up to allow for the telephone handset in the telephone condition).

An experimenter monitoring the conversations instructed the Wizards to ask the subjects to repeat an utterance during the course of the experiment when it was especially long, disfluent, or complex. These utterances by the Wizards, called "repetition requests" (RR), were usually of the form "Please repeat" for English[1]. This paper examines the clients' responses to these requests, and compares them with the initial utterances that provoked the requests.

Acoustic speech data was recorded on digital audio tapes using a SONY DAT deck, DTC-77ES. The acoustic tapes of the experiment sessions were transcribed, including notations for false starts; filled pauses such as "ah" and "uhum;" non-speech noises such as deep breaths or lip smacks; and simultaneous speech. The ten conversations comprised more than 12,600 words in over 1900 turns. There were 161 turns flanking requests for repetition; pre-RR utterances contained 2100 words; post-RR utterances contained 1580 words.

---

[1] Occasionally, the "Wizard" said "please speak slowly." These cases have been included in the analysis below when subjects in fact changed their utterances beyond merely slowing their speech. They have been excluded, however, from analyses involving speaking rate.

Attachment 1d: Copy of Document 1 from the Springer book

# Lecture Notes in Artificial Intelligence 1374

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

## Lecture Notes in Computer Science

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

10/7/99

F THE

198

LLINOIS
PAIGN

✓

Attachment 1d: Copy of Document 1 from the Springer book

Attachment 1d: Copy of Document 1 from the Springer book

# Lecture Notes in Artificial Intelligence 1374

Subseries of Lecture Notes in Computer Science
Edited by J. G. Carbonell and J. Siekmann

## Lecture Notes in Computer Science

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

✓

Springer
*Berlin*
*Heidelberg*
*New York*
*Barcelona*
*Budapest*
*Hong Kong*
*London*
*Milan*
*Paris*
*Santa Clara*
*Singapore*
*Tokyo*

Harry Bunt Robbert-Jan Beun
Tijn Borghuis (Eds.)

# Multimodal Human-Computer Communication

## Systems, Techniques, and Experiments

Springer

Attachment 1d: Copy of Document 1 from the Springer book

Series Editors
Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Harry Bunt
Tilburg University
Warandelaan 2, 5000 LE Tilburg, The Netherlands
E-mail: bunt@kub.nl

Robbert-Jan Beun
Center for Research on User-System Interaction (IPO)
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
E-mail: rjbeun@ipo.tue.nl

Tijn Borghuis
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
E-mail: tijn@win.tue.nl

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

**Multimodal human computer communication** : systems, techniques, and experiments / Harry Bunt ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1998
(Lecture notes in computer science ; Vol. 1374 : Lecture notes in artificial intelligence)
ISBN 3-540-64380-X

CR Subject Classification (1991): I.2, H.5.1-2, I.3.6, D.2.2, K.4.2

ISSN 0302-9743
ISBN 3-540-64380-X Springer-Verlag Berlin Heidelberg New York

Printed in Germany

Typesetting: Camera ready by author
SPIN 10631926 06/3142 – 5 4 3 2 1 0 Printed on acid-free paper

Attachment 1d: Copy of Document 1 from the Springer book

# Preface

This volume contains revised versions of seventeen selected papers from the First International Conference on Cooperative Multimodal Communication (CMC/95), held in Eindhoven, the Netherlands, in May 1995. This was the first conference in a series, of which the second one was held in Tilburg, The Netherlands, in January 1998. Three of these papers were presented by invited speakers; those by Mark Maybury, Bonnie Webber, and Kent Wittenburg. From the submitted papers that were accepted by the CMC/95 program committee, thirteen were selected for publication in this volume, after revision.

We thank the program committee for their excellent and timely feedback to authors of submitted papers, and at a later stage for advising on the contents of this volume and for providing additional suggestions for improving the selected contributions. The program committee consisted of Norman Badler, Harry Bunt, Jeroen Groenendijk, Walther von Hahn, Dieter Huber, Hans Kamp, John Lee, Joseph Mariani, Mark Maybury, Paul Mc Kevitt, Rob Nederpelt, Kees van Overveld, Ray Perrault, Donia Scott, Wolfgang Wahlster, Bonnie Webber, and Kent Wittenburg. We thank the Royal Dutch Academy of Sciences (KNAW) and the Organization for Cooperation among Universities in Brabant (SOBU) for their grants that made the conference possible.

January 1998

Harry Bunt
Robbert-Jan Beun
Tijn Borghuis

Attachment 1d: Copy of Document 1 from the Springer book

# Table of Contents

## Part I: Systems

## Part II: Techniques

## Part III: Experiments

# Multimodal Maps: An Agent-Based Approach

Adam Cheyer and Luc Julia

SRI International
333 Ravenswood Ave
Menlo Park, CA 94025 - USA

**Abstract.** In this paper, we discuss how multiple input modalities may be combined to produce more natural user interfaces. To illustrate this technique, we present a prototype map-based application for a travel planning domain. The application is distinguished by a synergistic combination of handwriting, gesture and speech modalities; access to existing data sources including the World Wide Web; and a mobile handheld interface. To implement the described application, a hierarchical distributed network of heterogeneous software agents was augmented by appropriate functionality for developing synergistic multimodal applications.

## 1 Introduction

As computer systems become more powerful and complex, efforts to make computer interfaces more simple and natural become increasingly important. Natural interfaces should be designed to facilitate communication in ways people are already accustomed to using. Such interfaces allow users to concentrate on the tasks they are trying to accomplish, not worry about what they must do to control the interface.

In this paper, we begin by discussing what input modalities humans are comfortable using when interacting with computers, and how these modalities should best be combined in order to produce natural interfaces. In Sect. 3, we present a prototype map-based application for the travel planning domain which uses a synergistic combination of several input modalities. Section 4 describes the agent-based approach we used to implement the application and the work on which it is based. In Sect. 5, we summarize our conclusions and future directions.

## 2 Natural Input

### 2.1 Input Modalities

Direct manipulation interface technologies are currently the most widely used techniques for creating user interfaces. Through the use of menus and a graphical user interface, users are presented with sets of discrete actions and the objects on which to perform them. Pointing devices such as a mouse facilitate selection of an object or action, and drag and drop techniques allow items to be moved or combined with other entities or actions.

With the addition of electronic pen devices, gestural drawings add a new dimension direct manipulation interfaces. Gestures allow users to communicate a surprisingly wide range of meaningful requests with a few simple strokes. Research has shown that multiple gestures can be combined to form dialog, with rules of temporal grouping overriding temporal sequencing (Rhyne, 1987). Gestural commands are particularly applicable to graphical or editing type tasks.

Direct manipulation interactions possess many desirable qualities: communication is generally fast and concise; input techniques are easy to learn and remember; the user has a good idea about what can be accomplished, as the visual presentation of the available actions is generally easily accessible. However, direct manipulation suffers from limitations when trying to access or describe entities which are not or can not be visualized by the user.

Limitations of direct manipulation style interfaces can be addressed by another interface technology, that of natural language interfaces. Natural language interfaces excel in describing entities that are not currently displayed on the monitor, in specifying temporal relations between entities or actions, and in identifying members of sets. These strengths are exactly the weaknesses of direct manipulation interfaces, and concurrently, the weaknesses of natural language interfaces (ambiguity, conceptual coverage, etc.) can be overcome by the strengths of direct manipulation.

Natural language content can be entered through different input modalities, including typing, handwriting, and speech. It is important to note that, while the same textual content can be provided by the three modalities, each modality has widely varying properties.

- Spoken language is the modality used first and foremost in human-human interactive problem solving (Cohen et al., 1990). Speech is an extremely fast medium, several times faster than typing or handwriting. In addition, speech input contains content that is not present in other forms of natural language input, such as prosidy, tone and characteristics of the speaker (age, sex, accent).
- Typing is the most common way of entering information into a computer, because it is reasonably fast, very accurate, and requires no computational resources.
- Handwriting has been shown to be useful for certain types of tasks, such as performing numerical calculations and manipulating names which are difficult to pronounce (Oviatt, 1994; Oviatt and Olson, 1994). Because of its relatively slow production rate, handwriting may induce users to produce different types of input than is generated by spoken language; abbreviations, symbols and non-grammatical patterns may be expected to be more prevalent amid written input.

### 2.2 Combination of Modalities

As noted in the previous section, direct manipulation and natural language seem to be very complementary modalities. It is therefore not surprising that a number of multimodal systems combine the two.

Notable among such systems is the Cohen's Shoptalk system (Cohen, 1992), a prototype manufacturing and decision-support system that aids in tasks such as quality assurance monitoring, and production scheduling. The natural language module of Shoptalk is based on the Chat-85 natural language system (Warren and Perreira, 1982) and is particularly good at handling time, tense, and temporal reasoning.

A number of systems have focused on combining the speed of speech with the reference provided by direct manipulation of a mouse pointer. Such systems include the XTRA system (Allegayer et al, 1989), CUBRICON (Neal and Shapiro, 1991), the PAC-Amodeus model (Nigay and Coutaz, 1993), and TAPAGE (Faure and Julia, 1994).

XTRA and CUBRICON are both systems that combine complex spoken input with mouse clicks, using several knowledge sources for reference identification. CUBRICON's domain is a map-based task, making it similar to the application developed in this paper. However, the two are different in that CUBRICON can only use direct manipulation to indicate a specific item, whereas our system produces a richer mixing of modalities by adding both gestural and written language as input modalities.

The PAC-Amodeus systems such as VoicePaint and Notebook allow the user to synergistically combine vocal or mouse-click commands when interacting with notes or graphical objects. However, due to the selected domains, the natural language input is very simple, generally of the style "Insert a note here".

TAPAGE is another system that allows true synergistic combination of spoken input with direct manipulation. Like PAC-Amodeus, TAPAGE's domain provides only simple linguistic input. However, TAPAGE uses a pen-based interface instead of a mouse, allowing gestural commands. TAPAGE, selected as a building block for our map application, will be described more in detail in Sect. 4.2.

Other interesting work regarding the simultaneous combination of handgestures and gaze can be found in Bolt (1980) and Koons, Sparrell and Thorisson (1993).

## 3 A Multimodal Map Application

In this section, we will describe a prototype map-based application for a travel planning domain. In order to provide the most natural user interface possible, the system permits the user to simultaneously combine direct manipulation, gestural drawings, handwritten, typed and spoken natural language. When designing the system, other criteria were considered as well:

# Contexts in dialogue

Tijn Borghuis

Department of Mathematics and Computing Science Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: tijn@win.tue.nl

**Abstract**

This paper shows how the modal type theory developed in [Borghuis 1994] can be used in formalizing communication. Based on the idea that the information states of the participants in a dialogue can be represented as a type theoretical context ([Ahn 1992]), the paper argues that in modal type theory an incremental representation of the content of the utterances in a dialogue can be brought together with a formal description of the effects of the pragmatic force of these utterances. To illustrate this I propose a formal procedure representing the update of the information state of a hearer by a declarative utterance of the speaker. This proposal combines existing work on discourse representation in type theory ([Ahn and Kolb 1990]) with existing work on epistemic pragmatics ([Thijsse 1992]) in the framework of a modal typed $\lambda$-calculus.

**Key words:** type theoretical contexts, Discourse Representation Theory, epistemic pragmatics, modal type theory.

## 1 Introduction: contexts as growing information states

In [Ahn 1992] a type theoretical approach to the formalization of communication is proposed. This approach, which is one of the fundaments of the DenK-project ([Bunt et al. 1995]), has as its central idea that the information state of an agent (animate or inanimate) can be modelled by a type theoretical context. In this view, the assertions that make up an agent's information state are represented as statements of the form $A : B$, where the type ($B$) of a statement corresponds to an assertion of the agent and the term ($A$) inhabiting the type corresponds to the 'justification' or 'evidence' the agent has for this assertion. In general, the information state of an agent will not contain a complete (or even accurate) description of the world: an agent may be uncertain about some propositions and unaware of others. Since the information state is incomplete, it may 'grow' as the agent learns more about the world. This growth can be modelled by appending statements representing the new information to the context representing the agent's information state. One source of growth is communication between agents, and [Ahn 1992] sketches a perspective under which dialogue can be viewed type theoretically as an exchange of information between (growing) contexts.

Against the background of Ahn's ideas, I construct a procedure for a particular instance of information growth in dialogue: the 'update' of the information state of a hearer-agent by a declarative utterance of a speaker-agent. The object of this exercise is to show that in an extension of type theory ('modal' type theory) two aspects of the formalization of communication that are usually studied separately can be brought together: the incremental growth of the hearer's information state by the content of the speaker's dialogue contributions, and the effect of the pragmatic 'force' of an utterance on the hearer's information state. Both aspects are treated by means of existing work, respectively the translation of Discourse Representation Theory to type theory in [Ahn and Kolb 1990], and the epistemic analysis of the Gricean maximes in [Thijsse 1992].

I formulate this update in a simple dialogue situation involving two agents, a speaker ($S$) and a hearer ($H$). Since the effect of a single utterance of the speaker is considered, the agents have fixed roles: the speaker speaks, and the hearer listens. The information states of speaker and hearer are represented as type theoretical contexts. Such a context contains declarations of all entities that the agent assumes to exist, and of all assertions (along with their proofs) that he holds about the world. It also contains statements declaring the 'vocabulary' (predicates, functions, sets) in which these assertions are formulated. Assuming that these statements denote concepts that are somehow related to words in the language, agents speaking the same language must share a considerable amount of this vocabulary to make communication possible.

When large discrepancies between the vocabularies of the dialogue participants exist, misunderstandings will arise. In this paper I want to abstract from such misunderstandings, and hence simply make the participants' information states before the dialogue (their *initial contexts*) isomorphic, by assuming that they contain the same vocabulary. Under this assumption, the initial context of agents can only differ in the *elements* and *proofs*[1]. This means that speaker and hearer may be able to prove different assertions about the world, and may have different justifications for the same assertion. Similarly, they may be familiar with different elements of a given set ('individuals'), or differ in the sets that are inhabited for them.

## 2 Discourse Representation Theory in type theory

The Discourse Representation Theory (DRT) of Hans Kamp ([Kamp 1981]) is a formal method for constructing representations for texts (sequences of sentences) in three steps. Starting from the sentences in the discourse a 'Discourse Representation Structure' (DRS) is generated, processing them 'from left to right' by means of 'DRS-construction rules'. These structures are then interpreted in a model through a truthful embedding. I shall not go into the construction nor into the embedding of DRSs, since our main concern is the relation of (already constructed) DRSs to type theory. In [Ahn and Kolb 1990] a formal translation is given from DRSs into type theoretical contexts: each DRS corresponds to a 'segment'. Using this translation, the growth of the information state of an agent interpreting a text can be modelled by the extension of the context representing the agent's information state with the segment representing the text.

Ahn and Kolb do not give a direct translation of the two-dimensional representations into type theoretical contexts. They use an intermediate sequential format in which *DRS*s are written in the following form: $r_1, \ldots, r_n, E_1, \ldots, E_m$, where $(r_1, \ldots, r_n)$ are the discourse referents and the 'entries' $E_1, \ldots, E_m$ are of one of the following three forms:

---

[1] In type theory one can distinguish syntactically between vocabulary on the one hand and proofs and elements on the other.

- atomic condition, n-ary predicate applied to a number of discourse referents,
- a complex condition $D_1 \Rightarrow D_2$, where $D_1$ and $D_2$ are DRSs,
- a link $[R = N]$ or $[R = R']$, where $R$ and $R'$ are discourse referents, and $N$ is a name in the model.

Given such a sequential representation of a *DRS*, Ahn and Kolb propose the following translation of DRSs to type theoretical contexts: a sequence of the general form $r_1, \ldots, r_n$, $E_1, \ldots, E_m$ translates to a 'segment' of the general form $r1 : entity, \ldots, r_n : entity$, $y_1 : E_1, \ldots, y_m : E_m$. The discourse referents are translated directly into variables. This is in line with the intuition that set variables act as 'pointers', they make an object of a certain type available to the reasoner. Since *DRT* has no typing (properties are attributed to the referents via predication), all discourse referents are given the same (neutral) type 'entity'.

Entries are assertions, and as such translated as terms of type *Prop*. They get a fresh variable $(y_1, \ldots, y_m)$ assigned as their proof term; the entries represent the content of the discourse, not its justification. The three kinds of entries are accommodated type theoretically as follows. Atomic conditions are an $n$-ary predicate applied to a number of referents. These are translated to statements $P(r_1, \ldots, r_n) : Prop$. Complex conditions are of the form $D_1 \Rightarrow D_2$. Roughly speaking, they are translated as a (series of) $\Pi$-abstraction(s) connecting $D_1$ to (part of) $D_2$. I illustrate this by means of the infamous donkey sentence 'Every farmer who owns a donkey beats it'. For this sentence, the segments corresponding to $D_1$ and $D_2$ are

$D_1 :\ u : entity, v : entity, p_1 : farmer(u), p_2 : donkey(v), p_3 : owns(u, v)$
$D_2 :\ u : entity, v : entity, p_1 : farmer(u), p_2 : donkey(v), p_3 : owns(u, v), p_4 : beats(u, v)$.

The sequence $u, v, (farmer(u), donkey(v), owns(u, v) \Rightarrow beats(u, v))$ $(D_1 \Rightarrow D_2)$ is translated into the statement $z : (\Pi u : entity.\Pi v : entity.\Pi p_1 : farmer(u).\Pi p_2 : donkey(v).\Pi p_3 : owns(u, v).beats(u, v))$, where the $\Pi$s abstract over the elements of the segment corresponding to $D_1$, and the body of the abstraction is the $D_2$-segment minus the statements that are also in the $D_1$-segment (and $z$ is a fresh variable). This abstraction is the proof theoretical reflection of the semantical idea that $D_1 \Rightarrow D_2$ turns any assignment satisfying $D_1$ into an assignment satisfying $D_2$. An interpreter who already has entities $(x, y)$ in his context as well as proof that these entities are respectively a farmer and a donkey $(p_5 : farmer(x), p_6 : donkey(y))$ and that $x$ owns $y$ $(p_7 : owns(x, y))$, can derive a term $z(x, y, p_5, p_6, p_7)$ proving '$x$ beats $y$' $(beats(x, y))$ by applying all this information to the type theoretical translation of $D_1 \Rightarrow D_2$. Links are expressions of the form $R = R'$ or $R = N$, which 'link' a discourse referent $R$ to another discourse referent $(R')$ or a name in the model $(N)$. They can be expressed type theoretically as $y : (R = R')$ or $y : (R = N)^2$.

## 3 Epistemic pragmatics

Reasoning about information states of (other) agents plays an important role in communication. For instance, in an information dialogue it is not cooperative to ask your dialogue partner something you already know, or to ask him a question you know he cannot answer.

---

[2] Where '=' abbreviates the Leibniz-equality that is definable in the type system used here.

A famous attempt to codify 'cooperative' behaviour in dialogue was made by Grice (see for instance [Grice 1989]). He begins his top-down development of dialogue behaviour rules by stating the

> **Cooperation principle** Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

Starting from this principle, Grice discerns four categories of rules for dialogue behaviour ('maxims'), each characterized by a 'super maxim' of which the most important one is:

> **Quality** Try to make your contribution one that is true

This general advice is then specified further in two maxims:

> **Belief** Do not say what you believe to be false.

> **Evidence** Do not say something for which you lack sufficient evidence.

In [Thijsse 1992], the 'epistemic force' that is attributed to (declarative) utterances through the quality maxims is analyzed in terms of epistemic/doxastic logic; propositional logic extended with modal operators '$K_x$', signifying 'agent $x$ knows that ...', and '$B_x$', meaning 'agent $x$ believes that ...'. This analysis results in the following proposal for an 'utterance rule'.

$$UTT \quad x : `\varphi' \Rightarrow B_x K_x \varphi.$$

If an agent ($x$) utters the proposition $\varphi$ ($x$ :'$\varphi$') he should believe to know that $\varphi$, $\varphi$ should be a true justified belief of his. An important benchmark in the epistemic analysis of the quality maxims are Moore's paradoxes (cf. [Moore 1912]), sentences about self-belief of the kind

(1) $p$, but I do not believe that $p$: $p \wedge \neg B_i p$

(2) $p$, but I believe that not $p$: $p \wedge B_i \neg p$

The puzzling thing about these sentences is that although they are logically consistent (the logical translations given above have verifying models), they are absurd to utter. In [Hintikka 1962] a similar example involving self-knowledge is given

(3) $p$, but I do not know whether $p$: $p \wedge \neg K_i p \wedge \neg K_i \neg p$.

The peculiarity of these 'Moore-sentences' can be formally demonstrated after the application of *UTT* to their logical translations: the resulting formulas are inconsistent in epistemic/doxastic logic.

Moore-sentences are not only strange to utter, they are also strange to hear. The analysis of the epistemic force of utterances should account for this in terms of the effects of an utterance on the information state of the hearer. In general a hearer need not be convinced of what the speaker says, but it seems reasonable to assume that the hearer is convinced that the speaker is convinced of what he says. Thijsse calls this effect 'epistemic transfer', and he extends his proposal accordingly with the following rule describing this effect of uttering a proposition ($\varphi$) by the speaker ($x$) on the hearer ($y$):

$$\textit{epistemic transfer}\ x : `\varphi\text{'} \ \Rightarrow\ B_y K_y B_x K_x \varphi$$

The combination of modal operators in front of $\varphi$ shows that the hearer is as sure of the utterance of the speaker, $B_H K_H(B_S K_S \varphi)$, as the speaker is of the proposition he utters, $B_S K_S(\varphi)$. In other words, the rule *UTT* is available to the hearer and is internalized by him. Under *epistemic transfer* the logical translations of the Moore-sentences are again inconsistent in the epistemic/doxastic logic, showing that their utterance is also strange from the hearers' point of view.

The epistemic/doxastic system used by Thijsse is a combination of the logic $\mathbf{KT4}_{(m)}$ for the knowledge-operators of the agents and the logic $\mathbf{KD4}_{(m)}$ for their belief-operators[3]. This means that in the resulting system the principles **K** and **4** hold for both knowledge and belief:

$$\mathbf{K}\ \ K_a(\varphi \supset \psi) \supset (K_a\varphi \supset K_a\psi),\ B_a(\varphi \supset \psi) \supset (B_a\varphi \supset B_a\psi)$$

$$\mathbf{4}\ \ K_a\varphi \supset K_aK_a\varphi,\ B_a\varphi \supset B_aB_a\varphi$$

**K** states that every agent ($a$) knows the logical consequences of his knowledge and believes the logical consequences of his beliefs. **4** is the principle of 'positive introspection': if an agent knows something, $K_a\varphi$, he knows that he knows it, $K_aK_a\varphi$. Similarly, if he believes something, he believes that he believes it. The difference between knowledge and belief is reflected in the following two axioms:

$$\mathbf{T}\ \ K_a\varphi \supset \varphi \qquad \mathbf{D}\ \ B_a\varphi \supset \neg B_a\neg\varphi$$

**T** expresses the 'veracity' of knowledge: if an agent knows something ($K_a\varphi$), it has to be the case ($\varphi$). This is too strong for belief, since beliefs can be mistaken. Hence the weaker principle **D**, stating that belief must be consistent: an agent cannot believe a proposition ($B_a\varphi$) and its negation ($B_a\neg\varphi$) at the same time. The logics of knowledge and belief are related by the axiom

$$\mathbf{FK}\ \ K_a\varphi \supset B_a\varphi,$$

which expresses that an agent believes every proposition that he knows to hold.

## 4 The Modal Pure Type System $\lambda\Box PRED2$

The formal framework in which I want to combine discourse representation and epistemic pragmatics is the Modal Pure Type System (MPTS) $\lambda\Box PRED2$. This system is a modal extension of the well-known Pure Type System (PTS) $\lambda PRED2$ in [Barendregt 1992], which corresponds closely to second order intuitionistic predicate logic (see [Geuvers 1993]).

Essentially, the difference between PTSs and MPTSs is that in MPTSs information (propositions) can be marked with operators (modalities) indicating what kind of information it is ('knowledge', 'belief', etc.), and that MPTSs allow additional structure in contexts. In MPTSs, one can, at any moment, create an additional separate part of the context (called 'subordinate context', marked by '⧈ ') into which only information of a certain kind may be transferred, for instance propositions which are registered in the context as 'known by

---

[3] These are multi-agent versions of basic modal logics to which a formal introduction can be found in [Chellas 1980].)

agent B'. In this subordinate context one can then reason with the information under the usual type theoretical inference rules to draw new conclusions. These conclusions can then be put back into the original context prefixed with the appropriate label ('known by B'), after which the subordinate context is removed. Using subordinate contexts, the type theoretical representation of the information state of an agent can be temporarily 'partioned' according to (syntactic) criteria determined by the rules for transferring information from a context to its subordinate context, the so-called 'import' and 'export rules'.

Since MPTSs are a well-behaved extension of PTSs, see [Borghuis 1994], the work of [Ahn and Kolb 1990] on DRT remains valid in $\lambda\Box PRED2$. The epistemic/doxastic system used in [Thijsse 1992] can be interpreted in $\lambda\Box PRED2$ if the following modal rules[4] are adapted. For knowledge:

$$(K\,import)\ \frac{G\ \vdash\ M : K_a\varphi : Prop}{G \boxdot^K_a \varepsilon\ \vdash\ \check{k}^K_a M : \varphi} \qquad (K\,export)\ \frac{G \boxdot^K_a \varepsilon\ \vdash\ M : \varphi : Prop}{G\ \vdash\ \hat{k}^K_a M : K_a\varphi}$$

$$(4\,import)\ \frac{G\ \vdash\ M : K_a\varphi : Prop}{G \boxdot^K_a \varepsilon\ \vdash\ \check{4}^K_a M : K_a\varphi} \qquad (T\,export)\ \frac{G \boxdot^K_a \varepsilon\ \vdash\ M : \varphi : Prop}{G\ \vdash\ \hat{t}^K_a M : \varphi}$$

For belief:

$$(K\,import)\ \frac{G\ \vdash\ M : B_a\varphi : Prop}{G \boxdot^B_a \varepsilon\ \vdash\ \check{k}^B_a M : \varphi} \qquad (K\,export)\ \frac{G \boxdot^B_a \varepsilon\ \vdash\ M : \varphi : Prop}{G\ \vdash\ \hat{k}^B_a M : B_a\varphi}$$

$$(4\,import)\ \frac{G\ \vdash\ M : B_a\varphi : Prop}{G \boxdot^B_a \varepsilon\ \vdash\ \check{4}^B_a M : K_a\varphi} \qquad (D\,export)\ \frac{G \boxdot^B_a \varepsilon\ \vdash\ M : \varphi : Prop}{G\ \vdash\ \hat{d}^B_a M : \neg B_a \neg\varphi}$$

These derivations rules in $\lambda\Box PRED2$ correspond to the axioms of the same name in the previous section. The axiom (**FK**) relating knowledge and belief corresponds to the rule:

$$(FK\,import)\ \frac{G\ \vdash\ M : K_a\varphi : Prop}{G \boxdot^B_a \varepsilon\ \vdash\ \check{k}^{(K,B)}_a M : \varphi}$$

Since epistemic/doxastic logic in which *UTT* and *epistemic transfer* 'live' can be accommodated in an MPTS, the derivations made by speaker and hearer based on these rules have a counterpart in modal type theory. What remains to be done is incorporating the 'modalization' of uttered propositions prescribed by these pragmatic rules in a procedure for adding type theoretical representations of utterances to the context of the hearer.

## 5 Adding declarative utterances

In this section, the ingredients presented separately above are combined into a procedure for adding declarative utterances to the information state of the hearer.

Starting from a declarative utterance of the speaker, a type theoretical representation of its content can be obtained by taking what the speaker says (the sentences used) to be a discourse. For this discourse a *DRS* can be constructed, which is turned into a segment,

---

[4] For a complete specification of MPTSs, and formal description of their relation to modal logics the reader is referred to [Borghuis 1994].

$r_1 : entity, \ldots, r_n : entity, y_1 : E_1, \ldots, y_m : E_m$, via the 'Ahn and Kolb-translation'. Rather than adding this segment directly to the context of the hearer, I propose to add it in the 'decorated' form $r_1^S : entity, \ldots, r_n^S : entity, y_1 : B_H K_H B_S K_S E_1, \ldots, y_m : B_H K_H B_S K_S E_m$. The discourse referents $r_1, \ldots, r_n$ are marked with the agent index of the speaker, to signify that the context of the hearer was extended with these referents to accommodate an utterance of the speaker. Since these referents are created on account of the speaker, the hearer should be allowed to use them in reasoning about knowledge or beliefs of the speaker. The entries, which represent the propositional content of the utterance, are prefixed with the modality $B_H K_H B_S K_S$ prescribed by the epistemic transfer rule to account for the epistemic effect of the utterance on the hearer. In the general format of the previous section the rule for adding an utterance '$U$' of agent $(a)$ to the context $(\Gamma_b)$ of another agent $(b)$ looks as follows:

> *AddUtt*
> $a : \text{'}U\text{'} \Rightarrow \Gamma_b, r_1^a : entity, \ldots, r_n^a : entity, y_1 : B_b K_b B_a K_a E_1, \ldots, y_m : B_b K_b B_a K_a E_m$
> where $r_1 : entity, \ldots, r_n : entity, y_1 : E_1, \ldots, y_m : E_m$ is a type theoretical representation of the discourse $U$, and $r_1, \ldots, r_n, y_1, \ldots, y_m$ are fresh variables w.r.t. $\Gamma_b$.

To see whether the *AddUtt*-rule makes any sense, I start by checking a simple example with respect to the inferences the hearer can make using the information he gets by adding an utterance of the speaker. Suppose that the hearer $(H)$ is aware of the 'donkey-ownership rule', which says that every farmer who owns a donkey beats it $\Gamma_H \equiv \Gamma, z : (\Pi u : entity.\Pi v : entity.\Pi p_1 : farmer(u).\Pi p_2 : donkey(v).\Pi p_3 : owns(u, v).beats(u, v))$, and that the speaker $(S)$ utters the sentence:

(4) Pedro is a farmer, Jerry is a donkey, and Pedro owns Jerry.

Under *AddUtt* the context of the hearer will be extended with a decorated version of the segment corresponding to (4), $r_1 : entity, r_2 : entity, y_1 : (Pedro = r_1^S), y_2 : (Jerry = r_2^S), y_3 : owns(r_1, r_2)$, and become:

$$\Gamma_H \equiv \Gamma, z : (\Pi u : entity.\Pi v : entity.\Pi p_1 : farmer(u).\Pi p_2 : donkey(v).\Pi p_3 : owns(u, v)$$
$$.beats(u, v)), r_1^S : entity, r_2^S : entity, y_1 : B_H K_H B_S K_S (Pedro = r_1^S),$$
$$y_2 : B_H K_H B_S K_S (Jerry = r_2^S), y_3 : B_H K_H B_S K_S owns(r_1^S, r_2^S).$$

On this context, the hearer cannot in any way derive that Pedro beats Jerry: he cannot conclude that he believes this himself, since he is not convinced of the information provided by the speaker. It is also impossible for the hearer to prove that the speaker believes that Pedro beats Jerry, since the context contains no evidence that the speaker is aware of the donkey-ownership rule. Technically, the speaker-modality $B_S$ in front of the entries blocks all applications of the general 'donkey-ownership rule' known by the hearer to the information about Pedro and Jerry provided by the speaker. Hence in this example the *AddUtt*-rule seems cautious enough.

However, the rule should not be too cautious to allow the hearer to derive the peculiarity of the utterance of a Moore-sentence by the speaker. Moreover, using the greater expressivity of the type theoretical 'DRT-language' over propositional logic, it should be possible to take into account that the peculiarity of the utterance of a Moore-sentence may depend on a previous utterance in the dialogue. If the speaker were to utter

(5) Every farmer who owns a donkey beats it, but I don't believe that Pedro beats Jerry.

in isolation, the hearer would not be able to judge its utterance by the speaker as inconsistent: he lacks the information that the speaker is convinced that Pedro is a farmer and Jerry a donkey that is owned by Pedro. However, if the speaker were to utter (5) after an earlier utterance of (4), the hearer should be able to judge this combination of utterances inconsistent.

Before testing *AddUtt* on this example, it should be noted that the epistemic transfer rule in [Thijsse 1992] was not intended for epistemic predicate logic, and the fragment of *DRT* covered so far does not have a construction rule for intensional verbs like 'to believe' or 'to know'. As in the logical translation of these sentences, the intensional verbs are represented as modal operators, i.e. in a segment representing 'I believe that $\varphi$' the entries representing the content of $\varphi$ will be prefixed with the modal operator $B_i$ (entries are now formulas in modal predicate logic).

The formalization of the example starts in the situation where the current context of the hearer is '$\Gamma$' and the speaker has just uttered (4). In the same way as above, applying *UttAdd* to (4) extends the context of the hearer to:

$$\begin{aligned}
&\Gamma_H \equiv \Gamma, r_1^S : entity, r_2^S : entity,\\
&y_1 : B_H K_H B_S K_S(Pedro = r_1^S), y_2 : B_H K_H B_S K_S(Jerry = r_2^S),\\
&y_3 : B_H K_H B_S K_S farmer(r_1^S), y_4 : B_H K_H B_S K_S donkey(r_2^S),\\
&y_5 : B_H K_H B_S K_S B_S owns(r_1^S, r_2^S)
\end{aligned}$$

Assuming that the content of the dialogue between the utterance of (4) and (5) is represented in the hearer's context by the segment $\Gamma'$, the next relevant moment in the dialogue is where the speaker utters (5), which corresponds to the segment $r_3^S : entity, r_4^S : entity, y_8 : B_H K_H B_S K_S(Pedro = r_3^S), y_9 : B_H K_H B_S K_S(Jerry = r_4^S), y_6 : (\Pi u : entity.\Pi v : entity.\Pi p_1 : farmer(u).\Pi p_2 : donkey(v).\Pi p_3 : owns(u, v).beats(u, v)), y_7 : \neg B_S beats(r_1, r_2)$. Adding this segment under *AddUtt* results in the hearer context:

$$\begin{aligned}
&\Gamma_H \equiv \Gamma, r_1^S : entity, r_2^S : entity,\\
&y_1 : B_H K_H B_S K_S(Pedro = r_1^S), y_2 : B_H K_H B_S K_S(Jerry = r_2^S),\\
&y_3 : B_H K_H B_S K_S K_H farmer(r_1^S), y_4 : B_H K_H B_S K_S donkey(r_2^S),\\
&y_5 : B_H K_H B_S K_S B_S owns(r_1^S, r_2^S), \Gamma',\\
&r_3^S : entity, r_4^S : entity, y_8 : B_H K_H B_S K_S(Pedro = r_3^S), y_9 : B_H K_H B_S K_S(Jerry = r_4^S),\\
&y_6 : B_H K_H B_S K_S(\Pi u : entity.\Pi v : entity.\Pi p_1 : farmer(u).\Pi p_2 : donkey(v).\Pi p_3 :\\
&owns(u, v).beats(u, v)), y_7 : B_H K_H B_S K_S \neg B_S beats(r_3^S, r_4^S).
\end{aligned}$$

Note that in this context there are *two* discourse referents for Pedro ($r_1^S, r_3^S$) and for Jerry ($r_2^S, r_4^S$), where $r_1^S$ and $r_2^S$ were in troduced by adding uttterance (4) and $r_3^S$ and $r_4^S$ by adding (5). Since under *AddUtt* every utterance of the speaker is represented type theoretically via a *DRS*, the hearer will have to add new referents to his context with every utterance of the speaker, even if 'conversationally' no new referents have been introduced. Assuming that names are set variables in the 'vocabulary' that is shared between the (initial) contexts of the dialogue participants (cf. section 2), referents linked to the same name can be identified 'across utterances' in $\lambda\Box PRED2$ by deriving the Leibniz Identity of the 'old' and the 'new' referent: from $y_1 : B_H K_H B_S K_S(Pedro = r_1^S)$ and $y_3 : B_H K_H B_S K_S(Pedro = r_3^S)$, a proof object ($M_1$) can be constructed for $B_H K_H B_S K_S(r_1^S = r_3^S)$. Similarly, $y_2 : B_H K_H B_S K_S(Jerry = r_2^S)$ and $y_4 : B_H K_H B_S K_S(Jerry = r_4^S)$ suffice to construct an inhabitant ($M_2$) of $B_H K_H B_S K_S(r_2^S = r_4^S)$. Since $M_1$ and $M_2$ prove that the old referents for Pedro and Jerry are identical to the new referents, the hearer can interpret the information provided by the speaker's utterance of (5) as applying to the old referents: $B_H K_H B_S K_S \neg B_s(beats(r_1^S, r_2^S))$. Substituting $r_1^S$ for $r_S^3$ and $r_S^2$ for $r_S^4$ simplifies the context $\Gamma_H$ to:

$\Gamma_H \equiv \Gamma, r_1^S : entity, r_2^S : entity,$
$y_1 : B_H K_H B_S K_S(Pedro = r_1^S), y_2 : B_H K_H B_S K_S(Jerry = r_2^S),$
$y_3 : B_H K_H B_S K_S K_H farmer(r_1^S), y_4 : B_H K_H B_S K_S donkey(r_2^S),$
$y_5 : B_H K_H B_S K_S B_S owns(r_1^S, r_2^S), y_6 : B_H K_H B_S K_S(\Pi u : entity.\Pi v : entity.$
$\Pi p_1 : farmer(u).\Pi p_2 : donkey(v).\Pi p_3 : owns(u, v).beats(u, v)),$
$y_7 : B_H K_H B_S K_S \neg B_S beats(r_1^S, r_2^S).$

On this context the inconsistency of uttering (5) after (4) can be derived in much the same way as for the utterance of (2) under *epistemic transfer* in epistemic/doxastic propositional logic. Since the derivation is both too long and too wide to reproduce in full, I show only the crucial middle part and use a few abbreviations.

1. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash r_1^S : entity$
2. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash r_2^S : entity$
3. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_3 : farmer(r_1^S)$
4. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_4 : donkey(r_2^S)$
5. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_5 : owns(r_1^S, r_2^S)$
6. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_7 : \neg B_S beats(r_1^S, r_2^S)$
7. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_6 : (\Pi u.\Pi v.\Pi p_1.\Pi p_2.\Pi p_3.beats(u, v))$
8. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_6 r_1^S : (\Pi v.\Pi p_1.\Pi p_2.\Pi p_3.beats(r_1^S, v))$
9. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_6 r_1^S r_2^S : (\Pi p_1.\Pi p_2.\Pi p_3.beats(r_1^S, r_2^S))$
10. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_6 r_1^S r_2^S M_3 : (\Pi p_2.\Pi p_3.beats(r_1^S, r_2^S))$
11. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_6 r_1^S r_2^S M_3 M_4 : (\Pi p_3.beats(r_1^S, r_2^S))$
12. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_6 r_1^S r_2^S M_3 M_4 M_5 : beats(r_1^S, r_2^S)$
13. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \vdash \hat{k}(M_6 r_1^S r_2^S M_3 M_4 M_5) : K_S beats(r_1^S, r_2^S)$
14. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash \check{4}(\hat{k}(M_6 r_1^S r_2^S M_3 M_4 M_5)) : K_S beats(r_1^S, r_2^S)$
15. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \boxed{\square}{}^B_S \varepsilon \vdash \check{f}(\check{4}(\hat{k}(M_6 r_1^S r_2^S M_3 M_4 M_5))) : beats(r_1^S, r_2^S)$
16. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash \hat{k}(\check{f}(\check{4}(\hat{k}(M_6 r_1^S r_2^S M_3 M_4 M_5)))) : B_S beats(r_1^S, r_2^S)$
17. $\Gamma_H \boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S \varepsilon \vdash M_7(\hat{k}(\check{f}(\check{4}(\hat{k}(M_6 r_1^S r_2^S M_3 M_4 M_5))))) : \bot$

In the beginning of the derivation each of the statements added to $\Gamma_H$ except $y_1 : B_H K_H B_S K_S (Pedro = r_1^S)$ and $y_2 : B_H K_H B_S K_S(Jerry = r_2^S)$, is brought to a '$\boxed{\square}{}^B_H \varepsilon \boxed{\square}{}^K_H \varepsilon \boxed{\square}{}^B_S \varepsilon \boxed{\square}{}^K_S$'-subordinate context by 4 subsequent applications of $K$-import[5]. In this way the modalities are stripped from the types and a situation arises in which there is proof that $r_1^S$ is a farmer, $r_2^S$ is a donkey, $r_1^S$ owns $r_2^S$, that the donkey-ownership rule holds and that the speaker does not believe that $r_1^S$ beats $r_2^S$ (lines 1-7)[6]. Hence the donkey-ownership rule can be used in combination with the information about Pedro and Jerry supplied by the speaker, to obtain a proof of $beats(r_1^S, r_2^S)$ (line 8-12). Since this is derived inside a categorical $K_S$-subordinate proof, it follows by positive introspection ($K_S\varphi \supset K_S K_S \varphi$) that the speaker knows that $r_1^S$

---

[5] $r_1^S$ and $r_2^S$ are brought to this context by means of a different rule ($transfer_3^s$), see [Borghuis 1994] section 6.2.

[6] The proof objects $M_3$-$M_7$ are abbreviations, where $M_i \equiv \check{k}_S^K(\check{k}_S^B(\check{k}_H^K(\check{k}_H^B y_i)))$ for $i \in \{3, 4, 5, 6, 7\}$.

beats $r_2^S$ (line 12-14). Knowledge implies belief (line 14-16), and so the hearer has proof in the $\boxdot{}_H^B \, \varepsilon \, \boxdot{}_H^K \, \varepsilon \, \boxdot{}_S^B \, \varepsilon \, \boxdot{}_S^K$-subordinate context that the speaker both believes and disbelieves that $r_1^S$ beats $r_2^S$, a contradiction.

¿From this contradiction a number of conclusions are derivable on the context $\Gamma_H$ representing the information state of the hearer, depending on the combination of $K$-, $D$-, and $T$-export rules used to finish the derivation. These conclusions range from $B_H K_H B_S K_S \bot$, the hearer is convinced that the speaker is convinced of a contradiction, through $B_H K_H B_S \bot$, $B_H B_S K_S \bot$, and $B_H \bot$ to $\bot$; the information state of the hearer is has become inconsistent. Hence the pragmatic force of the utterance captured in the epistemic/doxastic analysis of the maxim of quality can be expressed in the modal type system, in combination with an incremental representation of the effect of the content of that utterance on the information state of the hearer.

## 6 Concluding remarks

The update procedure presented in this paper is merely intended to indicate how two aspects of the formalization of communication that have been (and are being) studied separately can be brought together in a single formal framework. As such the procedure is too simple-minded in a number of respects, a few of which are discussed below.

First of all, the procedure only takes Grice's Quality maxim into account. The other maximes have also been studied in epistemic pragmatics, and their formalization seems to involve additional modalities. For instance, one of the maxims of Quantity ('Do not make your contribution more informative than is required') would translate to an utterance rule requiring that one should not present one's dialogue partner with information that already is 'mutually known' or 'mutually believed'. Technically, MPTSs are flexible enough to deal with the required modalities and their interactions. The problem is to give a comprehensive epistemic analysis of the dialogue situation one wants to formalize, since more specific considerations than those addressed by the Gricean axioms may come into play. For instance, in a dialogue where one of the participants is an expert on the topic of conversation and the other participant is not, the epistemic force of an sentence may also depend on *who* utters it.

Secondly the procedure does not make full use of the expressivity of type theory. For all its merits, *DRT* in the form used here has one important drawback: it is untyped. The universe of discourse is totally unstructured; all information about referents must be expressed via predication. If the discourse calls for the introduction of, say, a donkey, the translation will yield a segment containing the statements $r_i$ : *entity*, $y_j$ : *donkey*$(r_i)$, whereas type theoretically this could have been expressed more directly by means of the set-type 'donkey': $r_i$ : *donkey*. Using the expressivity of type theory, a more direct correspondence between type theoretical representation and syntactic structure of natural language sentences can be achieved, e.g. representing nouns by set-types and adjectives as predicates over these types.

Finally, the identification of discourse referents in the segment representing an utterance of the speaker with objects already present in the context of the hearer is far more complicated then it may appear from the examples in this paper. For the case of referents linked to the same (rigidly designating) name this identification is possible with standard type theoretical means. However, the speaker may also pick out a referent by means of a definite description for which the hearer has to find the appropriate referent already present in his context. This problem becomes even more difficult in the setting of a multi-modal dialogue where the speaker may make use of extra-lingual means (e.g. visual cues). These ways of 'anchoring' new information conveyed by the speaker in the information state of the hearer cannot be formalized entirely inside modal type theory as presented here.

In the DenK-project this problem is attacked by means of a multi-layered interpretation process of (user-) utterances involving several formalisms (using several sources of information) which subsequently resolve definite descriptions and ambiguities until a 'disambiguated' type theoretical segment representing the utterance of the speaker results (see [Bunt et al. 1995]). Adding this segment to the context of the hearer in a way similar to that described in this paper is the final step in this interpretation process.

## Acknowledgements

This research was carried out in the project "Dialogue handling and Knowledge transfer"(DenK) supported by the Cooperation Center Tilburg and Eindhoven Universities in the Netherlands.

## References

| | |
|---|---|
| [Ahn and Kolb 1990] | Ahn, R and Kolb, H.P., Discourse representation meets constructive mathematics. In *Papers from the second symposium on logic and language*, Kálmán, L. and Pólós, L. (eds.), Akadémia Kiadó, Budapest. |
| [Ahn 1992] | Ahn, René, A type-theoretical approach to communication, in *Think!* 1.1, ITK Tilburg University. |
| [Barendregt 1992] | Barendregt, Henk, Lambda calculi with types. In *Handbook of logic in computer science*, Abramsky, Gabbay and Maibaum (eds.), Oxford University Press, Oxford. |
| [Borghuis 1994] | Borghuis, Tijn, *Coming to Terms with Modal Logic: on the interpretation of modalities in typed λ-calculus* Ph.D.thesis Eindhoven University of Technology. |
| [Bunt et al. 1995] | Bunt, Harry, Ahn, René, Beun, Robbert-Jan, Borghuis, Tijn & Van Overveld, Kees, *Cooperative Multimodal Communication in the DenK Project*. This volume. |
| [Chellas 1980] | Chellas, Brian F., *Modal logic: an introduction*. Cambridge University Press, Cambridge. |
| [Geuvers 1993] | Geuvers, Herman, *Logics and type systems*. Ph.D. thesis, University of Nijmegen. |
| [Grice 1989] | Grice, Paul, *Studies in the way of words*. Harvard University Press, Cambridge. |
| [Hintikka 1962] | Hintikka, J., *Knowledge and belief. An introduction to the logic of the two notions*. Cornell University Presss, Ithaca (New York). |

[Kamp 1981] Kamp, H., A theory of truth and semantic representation. In: *Formal methods in the study of natural language*, Groenendijk, J.A.G. et al (eds.), Amsterdam.

[Moore 1912] Moore, G,E. *Ethics.* London 1912, reprinted by Oxford University Press, Oxford 1958.

[Thijsse 1992] Thijsse, Elias G.C., *Partial logic and knowledge representation*. Ph.D. thesis, Tilburg University, Eburon, Delft.

# Management of Non-Standard Devices for Multimodal User Interfaces under UNIX / X11

Patrick Bourdot, Mike Krus, Rachid Gherbi

LIMSI-CNRS, BP 133, 91403 Orsay cedex (France)
e-mail: emux@limsi.fr.

### Abstract

Multimodal user interfaces under UNIX / X11 rarely do the fusion of monomodal events on a temporal criterion. The main difficulties to achieve this are the representation and the dating of *non-standard events* (vocal, gesture, touch screen, eye movement,... ) by the X server. To begin with, we study a *modality server* which extends the control of X events to *non-standard events*. This server also enables the client applications to query the devices' states. Then we present the *multimodal widgets* which are in charge of the fusion of the monomodal events they receive. We explain why and how these widgets use the device state information.

**Keywords:** non-standard event, modality server, multimodal widget.

## 1 Introduction

One of the most promising results of the research on the multimodal user interfaces is their capacity to transform the man-machine communication of industrial applications which manipulate 2D or 3D virtual spaces. For example, it has been shown that a user can be more "productive" on a CAD program when keyboard interactions are replaced by vocal ones [11]. Hence a few research works are proposing to rethink these tools in terms of multimodal interactions [7].

However manipulating 2D or 3D virtual spaces requires a powerful graphic environment. Today, many of these applications are developed on UNIX workstations using X11 as a standard tool, sometimes with additional hardware graphical functions. In spite of the various functions it can realise, the X server only manages, in terms of input modalities, mouse and keyboard events.

In this paper, we first recall some definitions about multimodal systems. Then we explain the problems that the UNIX / X11 environment induces, and we make a quick analysis of some previous works. Rather than using a Real Time kernel with an UNIX extension, we propose a pragmatical architecture which enables to discard such a solution for most multimodal applications. We then expose our principle of a *modality server* and introduce the concept of a *multimodal widget* to encapsulate multimodal interactions. Finally, we present a first application of the *modality server* and we conclude in pointing multimodal interface issues.

## 2 Some Definitions

### 2.1 Multimedia and Multimodal Applications

To explain the differences between multimedia and multimodal applications, an author [5] has introduced a dimensional classification into three axes: "levels of abstraction", "fusion levels" and "temporal constraints".

A multimedia application uses low "levels of abstraction", because it does not do semantical processing of the data (from various media) to obtain meaning. At the opposite, a multimodal application has several "levels of abstraction" from the raw data to symbolic representations. These representations allow the application to have artificial reasoning and to improve the computer-human interaction. Concerning the "temporal constraints" axe, this author makes a distinction between "sequential" and "concurrent" user interfaces. The latter are allowed to receive (resp. produce) multiple input (resp. output) expressions at the same time, while the former can only manage one expression at a time. In the same way, it is possible to considere two types of user interfaces on the "fusion levels" axe. A user interface is "exclusive" if each expression is built from only one modality. At the opposite, a "synergetic" interface is allowed to receive (resp. produce) expressions built from several input (resp. output) modalities.

Without the "levels of abstraction" axe, the "fusion levels" and "temporal constraints" axes imply four cases of user interfaces:

a. A "sequential" and "exclusive" user interface is like any regular user interface, except it may have recognition or synthesis processes on different modalities.

b. A "concurrent" and "exclusive" user interface is possible with any multitasking environment. As in the previous case, it could also have some symbolic processes on input or output modalities.

c. A "sequential" and "synergetic" user interface builds expressions from a chronological interlacing of input or output modalities. This type of user interface is multimodal because it needs syntactical representations to construct the expressions (which is a minimal step of symbolic representations into the "levels of abstraction").

d. A "concurrent" and "synergetic" user interface builds expressions from a synchronous combination of several input or output modalities. As in the previous case, this type of user interface is always multimodal because it needs, at least, syntactical representations to construct the expressions.

Thus, the possibility to combine within expressions several modalities which are chronologically or synchronously processed is a very important characteristic of the multimodal applications. In the following we study the fusion of input modalities. An input modality generally produces monomodal events (the dual information, which is the states of devices, is also useful but we will speak about them only in the **6.2 section** when we introduce the notion of *multimodal widget*). It is these events (or dual states) that we have to combine. By hypothesis, a monomodal event can be the result of a recognition system associated to the modality's input device.

### 2.2 The Criteria for the Fusion of Monomodal Events

The fusion of monomodal events can be prepared at the lowest "level of abstraction". It is then performed with respect to integration criteria. Five of them are presented in [6]. We shall introduce them and then focus on a specific criterion.

The first one is the "logical (structural) complementary" of events. It allows, in some cases, to combine temporally distant events within the same command. The second one is the "data structure completeness". It can constitute a condition to move within the "abstraction levels". This completeness is also useful to reduce waiting events. The third one is the "dialogue contexts". It is used with the historical log of the interactions to resolve co-references between modalities (when a modality interaction cannot be correctly understood without events or states from others modalities) and to manage anaphora, ellipsis or deictic expressions. The fourth one is the "incompatibility of modalities". It allows to avoid the integration of modalities that cannot be used together.

Finally, the last criterion is the "temporal proximity". It is this one that we will study now.

## 3 The Problem

### 3.1 The Criterion of Temporal Proximity

The temporal proximity allows to simulate "concurrent" and "synergetic" multimodal user interfaces on monoprocessor workstations. It is also important for a physiological reason due to the interactions that a human operator can perform on a multimodal application. Combinations of modalities by a human operator have not the temporal exactitude and precision of a machine, which implies that input "concurrent" interactions would have no sense without a short time delay.

But the temporal proximity is mainly used as a criterion to combine different modalities. For instance, the fusion of input events proposed by [1] first determines the events which are produced within a short time range, and then combines some of them with respect to semantic criteria like those mentioned before. With this fusion process, the well-know multimodal command "put this here" associated with two graphical selections becomes possible. The temporal proximity filter allows to determine the co-references between the "this" and "here" vocal events, and the first and second selections respectively.

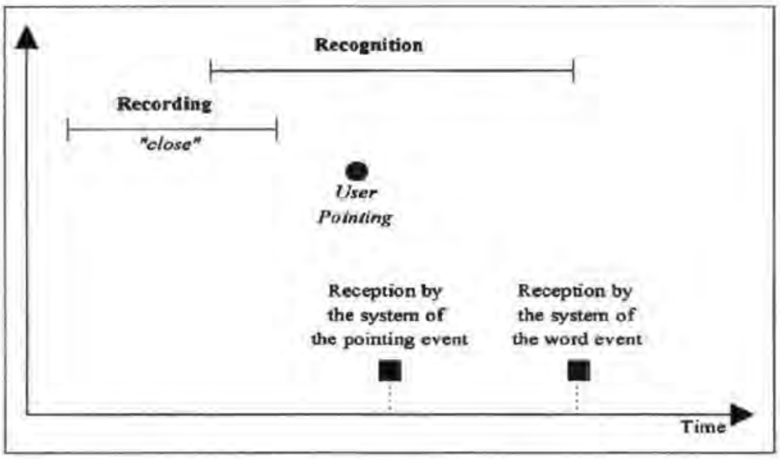


Figure 1: The problem of devices' response times with speech recognition (taken from [3]).

On the other hand, a same sequence of monomodal events can produce different multimodal expressions because semantic interpretations can depend on the time distribution of the events. But, as it is explained in [3], devices' response times may be so different that misinterpretations are possible when the system has to decide the merging of monomodal events (figure 1). That is the specific case of the devices associated to recognition systems (speech, gesture,...), because they need much more time to analyse an expression than any standard input devices (mouse, keyboard,...) or than any non-standard devices without recognition process (tactile screen, eye-tracker,...). Hence, the only solution to avoid undesirable actions on the objects of the multimodal application, is to know exactly the starting date and the duration of the recognition process of each monomodal expression.

All these examples show that a precise time stamping is necessary for the fusion of monomodal events. But the problem becomes difficult because the management of non-standard devices (such as: tactile screens, eye-trackers, vocal or gesture recognition systems) has to be done within an UNIX / X11 environment.

### 3.2 The UNIX / X11 Environment

An X server knows how to manage mouse and keyboard events perfectly. But in the case of *non-standard events*, one must be able to represent these events as well as giving a date in a manner that is coherent with X. The problem of representation is partially resolved by X itself as it offers the appropriate structures to add new types of events. On the other hand, precisely dating a *non-standard event* within X is more difficult, especially when the event is the result of a recognition process.

Indeed, in the case of recognition, we are interested in the date at which the event begins to be produced by the user. Unfortunately, the date that the X server would give to the event when putting it in the event queue could only be a later date. Besides, even if this date could be corrected in order to approach the desired date, this would not affect the position of this event within the event queue. Furthermore, X server does not allow an application to modify the order of the events in the queue. Finally, the date of the X server is a counter which is incremented "roughly" one thousand times a second. This date cannot be directly known and one must use events to find it out.

In this context, X only allows a late dating of the *non-standard events* which are the result of recognition. This increases the uncertainty of such event dates. To guaranty the validity of these dates we must use the internal clock of the workstation. But time representation for X and for the workstation is completely different. So one must provide a mechanism to convert the date from a format to the other.

### 3.3 Bibliographical Comments

Some authors seem to have realised multimodal systems with both graphics and vocal commands under UNIX / X11.

The most popular work is Xspeak [13], where an X extension enables the representation of vocal events. However, the multimodality supported by this system does not seem to use temporal proximity as a fusion criterion. The same remark seems to be true for Munix (Multimodal UNIX) [9], a project which aims to increase the ergonomic qualities of the UNIX operating system by partially substituting vocal commands to keyboard interaction. The GEORAL project offers graphical and vocal interaction with a geographical database [8], but this multimodal application combines vocal and graphics for output only. Finally in [12], the authors use vocal and mouse interactions as input for bilingual translation tool, but do not combine them since vocal modality is only used for translation queries. It suggests that a system of speech recognition should be considered as a server. In fact, we believe that this principle has to be extended to any modality producing *non-standard events*.

## 4 Co-operative approach

### 4.1 Architectural Principle

On monoprocessor multitasking workstation, the necessity of precise dating for *non-standard events* also raises the question of using an operating system based on a Real Time kernel. A high Real Time priority is useful for any modality's process which has to manage a recognition system. An other process with a higher priority is also necessary for the dating of *non-standard events*. However, a regular time sharing is sufficient for the X server and the multimodal application.

Unfortunately, there is not at this time any standard UNIX extension to a Real Time kernel. So we decided to use a UNIX with Real Time extensions (as UNIX System V R4.3). But according to [10], this type of operating system can be used in place of a Real Time kernel with a UNIX extension, only if the applications require a minimal processing time (including context swap) greater than 100 milliseconds.

Because of this processing time limit, it becomes necessary that the dating process of *non-standard events* has, with no restriction, the highest Real Time priority. It is the reason why we chose to use a distributed architecture (figure 2). To guarantee Real Time, we dedicate a "slave" processor for each modality which has a recognition system. On a "master" processor, a process with the highest priority manages the dating of the events recognised by the "slave" processors or sent by the non-standard devices without recognition systems. Furthermore, the "master" processor controls the multimodal interactions and executes the applications. Depending on the nature of the recognition systems, the "slave" processors may or may not be multitasking, while the "master" processor will always be.

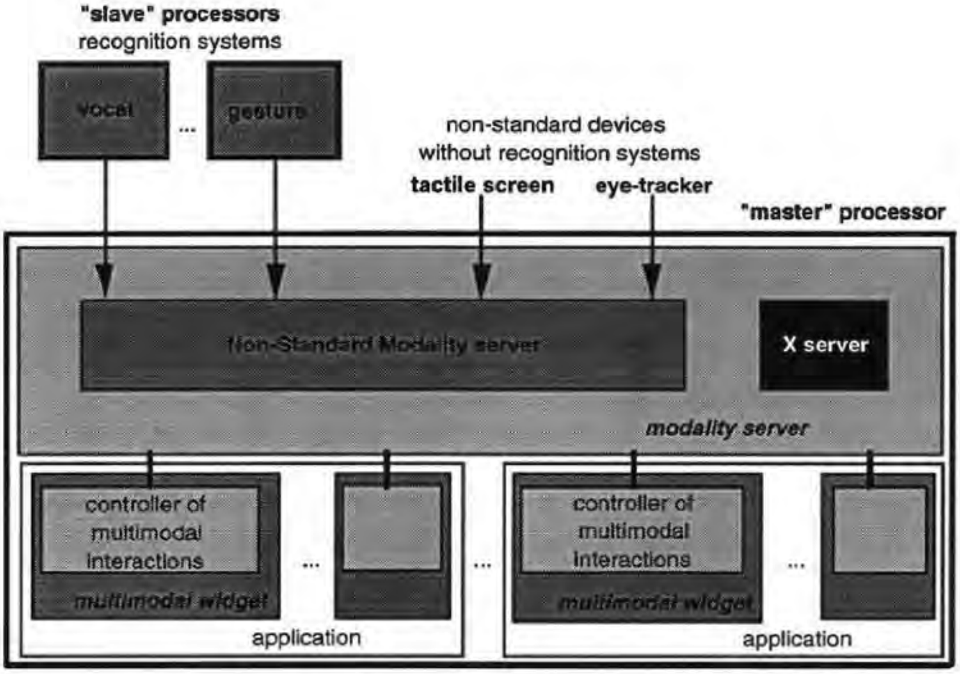


Figure 2: architecture for "concurrent" and "synergetic" multimodal user interfaces.

In practice, we simulate this architecture by a co-operation between machines. We have experimented two kinds of data transmission: the connection by serial port and by TCP / IP network. The first one is deterministic since its transfer time is known and constant, while the second one is faster but more sensitive to the load on the network.

### 4.2 Context and Situation

The fundamental aspect of this work lies mainly on the fact that we want to elaborate a generic multimodal kernel, which can be used in several and different applications. However, we propose pragmatical solutions guided by an experimental method. From this point of view, we do not presently invoke or study the dialog aspect of this kernel. We think that before doing such study we must resolve all problems encountered in event based interactions. Indeed, the principle problems concerning the representation and the dating of *non-standard events* have enormous importance.

Another important point that we claim is the ability of our system to rapidly allow the adjunction of other modalities (gestural, vision, drawing recognition) without completely reconsidering the system's architecture. Beside, we do not favour any modalities, in order to let this choice to the end-user or to the interface-designer. This also implies a certain universal kernel architecture.

# 5 Modality Server

## 5.1 General Principle

Our approach is based on a client-server concept between multiple X clients (figure 3). A *modality server* combines the X server with a particular X client. This X client acts as a Non-Standard Modality server (also known as NSM server) for all multimodal application clients of the X server. The NSM server is itself composed of several modules.

The main purpose of the TRANSPORT module of the NSM server is to monitor the input ports of the "master" machine. It is listening for any events sent by the "slave" machines or by the non-standard devices directly connected to the "master" machine (this is a non-busy wait). When such events occur, they are sent to the module associated with that modality. Depending on the devices, the result may be a set of one or more *non-standard events*. After that, the INTERFACE module of the NSM server tells the client which currently has the focus that a set of *non-standard events* has occurred, by sending him a "ClientMessage" event.

Unfortunately, this type of X event has a bandwidth which is insufficient for sending all information concerning a set of *non-standard events*. This can be handled by different methods of communication between process, like sockets, pipes, and so on. For data access rapidity, our present prototype of *modality server* is placing all the information in a shared memory segment and sending only its identification to the client. When the client receives a "ClientMessage" from a modality module, it reads the information and sends immediately after to this module (via the X server and the INTERFACE module) the order to release the memory segment.

## 5.2 Dating Non-Standard Events

But a "ClientMessage" is placed without a date in the X event queue. This represents a part of the general problem of dating *non-standard events* and of managing the queue of monomodal events. In the following, we will see how we resolve such a disadvantage.

As we already explained in the **3.2 section**, the only possibility to know the date of the X server is to create a regular X event. In this aim, a solution to give an X date to the "ClientMessage" of a modality module is to create a pseudo X event just after each "ClientMessage". But this dating is too imprecise to be used in the computation of the "real" dates of the *non-standard events* in the set. This is the reason why the NSM server includes a DATE CONTROLLER module.

When a set of *non-standard events* just arrives to a monomodality module, the DATE CONTROLLER asks the workstation clock to get the current time and converts it to the equivalent X date. For this conversion, a link must exist between X and the workstation times. This link can be a simple calibration of the workstation clock with the X date format. In our prototype of *modality server*, we make this calibration only when the NSM server is started.

Then the "real" times of the *non-standard events* of a set can be computed based on parameters particular to each modality (acquisition time, recognition time, data transfer time,...).

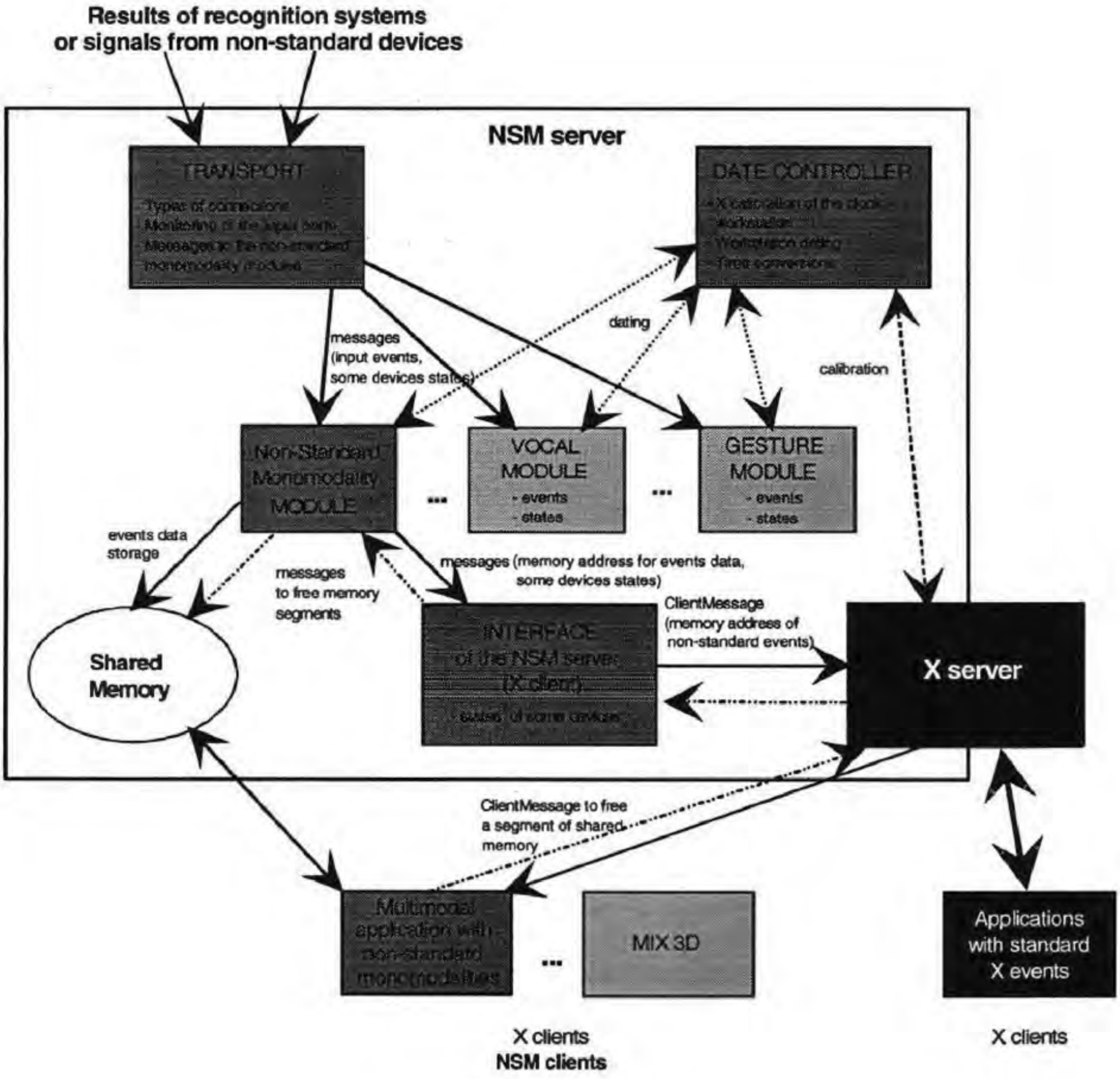


Figure 3: the *modality server.*

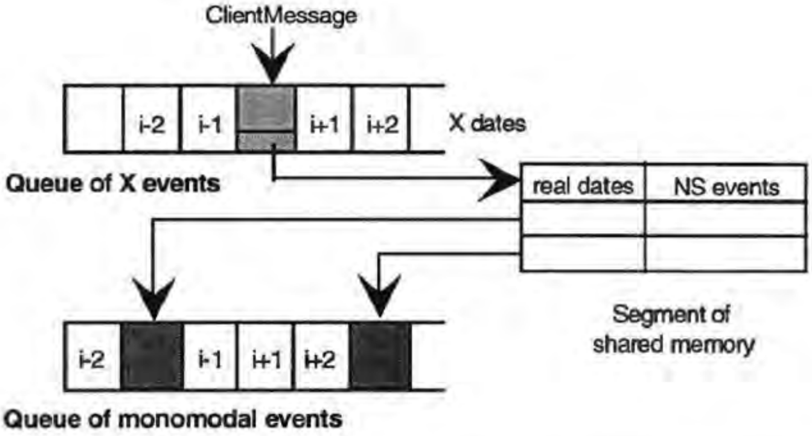


Figure 4: the management of the queue of monomodal events.

# 6 Multimodal Widget

## 6.1 The Management of the Multimodal Interactions

Our goal is to extend the X11 model to handle multimodal interactions. Most applications written for the UNIX / X11 environment use Motif widgets to represent familiar elements of the user interface. However the constraints on widgets, both for input and output, are too low-level for multimodal interactions. For this reason, [4] introduces the concept of "metawidgets". Such widgets are the links through which the application and the interface exchange conceptual information. But [4] only focused on multimodal presentation using alternate modalities. Our purpose is to use a *multimodal widget* for input, by having it operate the fusion of monomodal events from several modalities into a multimodal message.

This type of widget has a particular event manager to control the multimodal interactions (figure 2). Part of this management is to decide which monomodal events have to be combined to create multimodal ones. This process needs a specific queue where the monomodal events are ordered with respect to their "real" dates which have been computed by the DATE CONTROLLER. So a *multimodal widget* has to sort the monomodal events it receives. In order to do this, the *non-standard events* pointed by a single "ClientMessage" are read in the shared memory and inserted in the correct position, among other regular X events (figure 4).

But this fusion process must not combines monomodal events in any way. Each *multimodal widget* of an application needs a description of the multimodal interactions this widget supports, to know how monomodal events have to be combined. To validate this approach we are developing a *multimodal widget* prototype where its interactions are described by an Augmented Transition Network [14], according to the model defined in [1] to realise SPECIMEN (a multimodal interface specification tool).

## 6.2 The Use of the Device States

The TRANSPORT module has an other job which is to communicate to a modality module the state of its "slave" machine or the state of its non-standard device (when this one is directly connected to the "master" machine). The INTERFACE module finally transforms each state information into a property that any application which uses the corresponding modality can consult.

First of all, the state information is useful for a *multimodal widget* during the fusion process of the monomodal events. When a multimodal event is under construction, it is necessary to be sure that a recognition system is in a waiting state and is not creating new monomodal events. If this happens it could affect the queue of the monomodal events during the time interval which is scanned by the fusion process. In this case this process must wait until the end of the recognition to realise a relevant fusion of monomodal events.

Second, the state information is also useful to solve the "passive co-reference" [2]. It means that an event cannot be correctly understood by a system, if this system does not know the state of one or several devices at the date of this event. This problem appears when the frequency of the monomodal event production of a non-standard device is relatively higher than other modalities. In this case, we propose that the multimodal widgets could manage historical queues of states for each device causing this problem. To illustrate our purpose a good example is an interactive application which combines an eye-tracker device with a speech recognition system. When a vocal command is recognised, it is necessary to have an historical queue of the states of the eye-tracker, in order to find the direction pointed by the eyes when the operator gave the speech command.

## 7 Application

We have already used a part of the *modality server* to realise MIX3D (Multimodal Interactions in a X environment with a 3D virtual space). That is a 3D modelling program where interactions are "non-concurrent" combinations of vocal and graphical commands. A "slave" machine, with a Datavox card of VECSYS, is in charge of the vocal recognition. The "master" machine, a SiliconGraphics, manages the *modality server* and the application program. In this hardware and software context, we especially verify that our *modality server* does not create problems with local graphics libraries (such as GLX, the X extension of the graphic library of SiliconGraphics).

But this test application had mainly confirmed that a multimodal user interface can really reduce the cognitive stress of a CAD operator. For instance, the operator is able to focus his perception on the graphic work space, making the visual control for the menu and keyboard interactions optional. When speech allows the user hands to be free of the mouse interactions with the menus, it does not only increase the precision of his graphic task on 3D virtual objects. Just like it is when vocal commands replace keyboard interactions, a multimodal interaction with speech input more fundamentally avoids changes of working context for the user. On the other hand, most of the commands that a CAD operator has to do are already known by him during his graphic interactions. That is because the user generally plans his work. Even if a complete verbalisation of the commands is limited by the operator's tiredness, it appears clearly that the vocal input modality is a clever way to take advantage of this planning.

Another issue of multimodal interactions with speech input is the generalization of the "put this here" paradigm. For example, suppose the multimodal interaction "put the red door here", where a graphic selection of a rotation axe would be the co-reference of the "here" event. In this case, the "red" and "door" events make references to particular concepts. But we naturally cannot built menus for all concepts that a CAD operator could manipulate. It is clear, for such concepts, that speech input allows to avoid the heaviness of textual interaction

Figure 5: the multimodal MIX3D workstation.

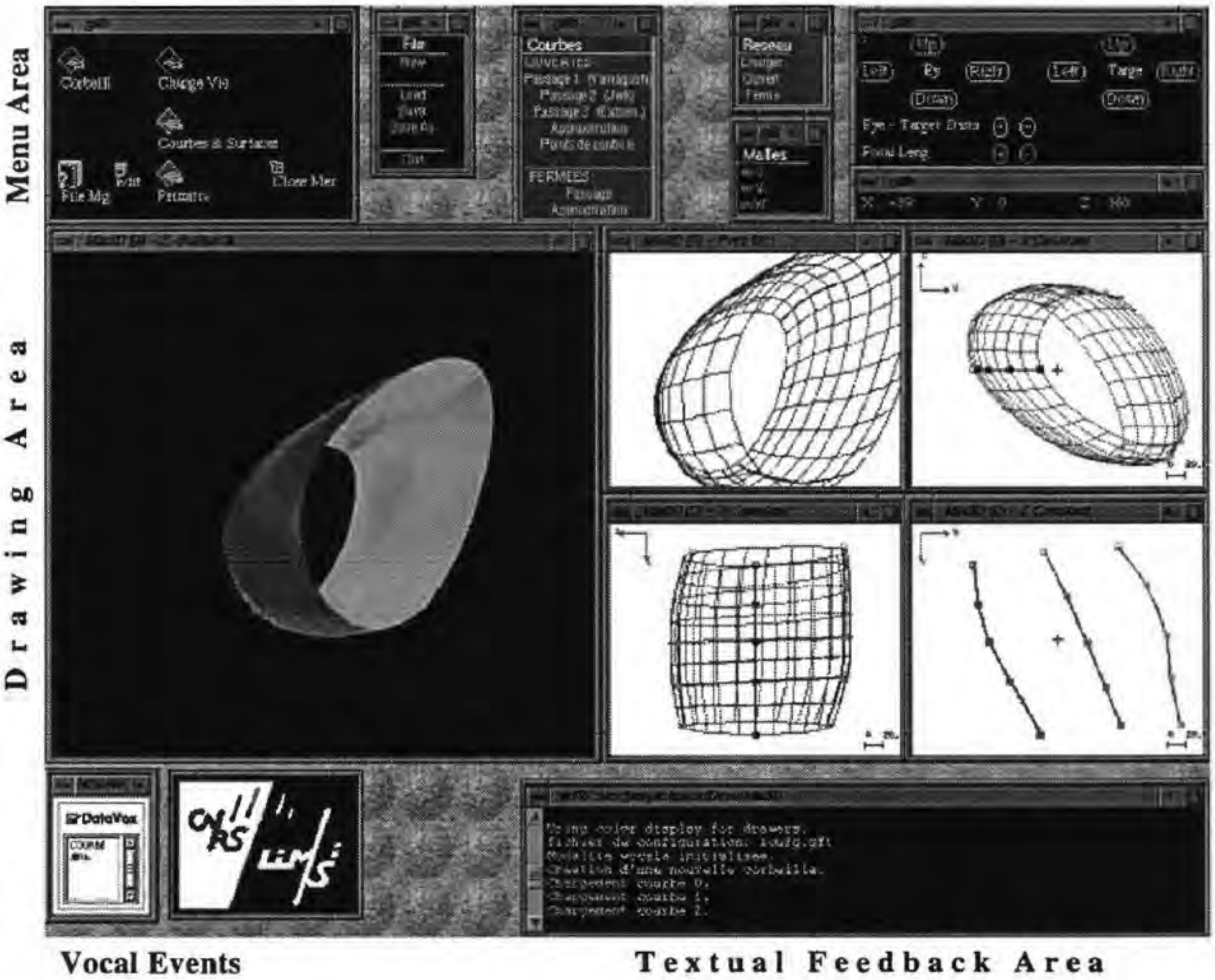


Figure 6: the multimodal MIX3D screen.

which is the only other alternative.

Apart from input modalities, MIX3D had also allowed to test the cooperation between output modalities. For instance, we use vocal output when a command modifies a 3D object without graphical feedback. Every vocal message of this type is associated to a textual message in a standard output window. If the combination of these two modalities has the same objective and could have partial redundant informations, these messages do not have the same use but could be complementary. The vocal output is an immediate and volatile control of the command's result. It avoids the user having to look for messages on standard output windows. The textual message serves as tracing purposes may be useful for controls later in the work session.

Finally, this test application has also suggested the association of some speech output messages with the main menu or keyboard input commands. Our purpose is not to create simple verbal help since it exists in textual form. Depending on the quality of the operator's interactions with the multimodal interface of an application, these speech output messages would teach the user (or rewind him) the correct vocal command which can be substituted to a menu or keyboard one.

## 8 Conclusion

Out of any industrial application context, we have already verified the reliability of the dating of the *non-standard events* made by our modality server. In the same way, we have also controlled its efficiency to deliver the device state information. The current phase of our work is to introduce *multimodal widgets* within MIX3D to realise a complete validation of our *modality server*. The next step will be to use the modularity of this server to connect MIX3D with the gesture recognition system which is designing at the LIMSI-CNRS laboratory.

## References

[1] Y. Bellik, D. Teil - *A multimodal dialogue controller for multimodal user interface management system, application: a multimodal window manager* - INTERCHI'93, Amsterdam; April 1993.

[2] Y. Bellik, S. Ferrari, F. Neel, D. Teil - *Communication multimodale: approche ascendante / application au domaine spatial* - Ecole SIC'94, Paris; April 1994.

[3] Y. Bellik, S. Ferrari, F. Neel, D. Teil - *Requirements for multimodal dialogue including vocal interaction* - ESCA Tutorial and Research Workshop on Spoken Dialogue Systems, Hanstholm (Denmark); May 1995.

[4] M.M. Blattner, E.P. Glinert, J.A. Jorge, G.R. Ormsby - *Metawidgets: Towards a theory of multimodal interface design* - COMPSAC'92, Chicago, IEEE Computer Society Press; September 1992.

[5] J. Coutaz - *Multimedia and multimodal user interfaces: a software engineering perspective* - International Workshop on Human Computer Interaction, St. Petersburg; 1992.

[6] J. Coutaz et al. - *Interfaces multimodales et architecture logicielle* - Compte rendu des ateliers des 4hmes journies sur l'inginierie des Interfaces Homme-Machine, IHM'92, Paris; December 1992.

[7] V. Gaildrat, N. Vigouroux, R. Caubet, G. Pirennou - *Conception d'une interface multimodale pour un modeleur diclaratif de schnes tridimensionnelles pour la synthhse d'images* - L'interface des mondes riels et virtuels, Informatique 93, Montpellier; March 1993.

[8] F. Gavignet, M. Guyomard, J. Siroux - *Mise en oeuvre d'une application multimodale orale et graphique: le projet GEORAL* - Actes des 2hmes journies nationales du GRECO-PRC Communication Homme-Machine, Toulouse; January 1991.

[9] P. Lefebvre, G. Duncan, F. Poirier - *Speaking with computer: a multimodal approach* - Eurospeech'93, Berlin; 1993.

[10] P. Le Foll - *Intigration des applications temps riel dans l'informatique de l'entreprise* - Siminaires de la confirence Real Time System, RTS'94, Paris; January 1994.

[11] G. L. Martin - *The utility of speech input in user-computer interfaces* - International Journal of Man-Machine Studies, V30; 1989.

[12] V. Normand, J. M. Lunati - *Diveloppement d'une application multimodale: une expirience* - Actes des 2hmes journies nationales du GRECO-PRC Communication Homme-Machine, Toulouse; January 1991.

[13] C. Schmandt, M. S. Ackerman, D. Hindus - *Augmenting a window system with speech input* - IEEE Computer, V11, N8; August 1990.

[14] T. Winograd - *Language as a cognitive process* - V1, Syntax, Addison Wesley; 1983.

# The Role of Multimodal Communication in Cooperation and Intention Recognition: The Case of Air Traffic Control

Marie-Christine Bressolle[1], Bernard Pavard[1], Marcel Leroux[2]

[1]ARAMIIHS 31 rue des Cosmonautes ZI du Palays
31077 TOULOUSE CEDEX FRANCE
e-mail: bressoll@cenatls.cena.dgac.fr, pavard@jazz.matra-espace.fr
[2]Centre d'Etudes de la Navigation Aerienne 7 avenue Edouard Belin
31055 TOULOUSE CEDEX FRANCE
e-mail: leroux@cenatls.cena.dgac.fr

**Abstract**

This study aims at the understanding of cooperative cognitive strategies used by air traffic controllers in simulated situations. In such situations, cooperation is mainly based on Intention Recognition processes. We have shown that in such close-to-reality situations, intention recognition is made possible because of multimodal communications which deeply involve external artifacts (strips, radar scope, radiotelephony). We will assume that the efficiency of communication is made possible only if the environment can allow intention recognition through non-verbal channels. Following this hypothesis, we will discuss the implications of such results on a general method for design based on assessment of the multimodal wealth of the working environment.

**Key words:** Intention recognition, non-verbal communication, distributed cognition, environmental resources.

## 1 Introduction

This study aims at the understanding of verbal and non-verbal modalities by which the agents cooperate, by focusing on informal communication mechanisms which could be modified or changed by the introduction of a new technological environment. As shown by numerous studies in the domain of CSCW, the study of cooperative processes between agents is essential to investigate or validate choices of new technologies or new modes of interaction between human agents. The design of new working environments brings up questions such as what information on the activities performed by other agents should be presented to a given agent (for example partially shared information, respecting confidentiality), or what are the operational procedures required to support implicit or explicit exchanges between agents, etc. For example, different approaches and their consequences on human coordination required in specific fields have to be examined (Dourish and Belloti, 1992). Understanding the cooperative dimension in work activities is complex because of the interaction between social and cognitive dimensions and of the nature of the processes involved in intention recognition which calls upon different modalities (writing, moving objects, pointing, etc.)

When examining different approches, focusing on the cooperative nature of working practices in different areas, it appears that understanding cooperation between human agents leads to consider collective cognition as socially distributed with respect to external artifacts used in practice. In the domain of London underground control, Heath and Luff (1991) found that controllers develop a practice of overhearing each other's conversations and overseeing each other's actions. This allows them to manage a flexible division of labour well adapted to solve difficulties. This flexibility seems to be dependant on the ability to manage implicit task allocation in the team. Some researchers in the domain of Air Traffic Control, (Hughes et al. 1992) take into account the artifacts used in such situations to highlight how division of labour is related to working practice using artifacts to organize activities within the team (for example, writing on strips is seen as support to cross-check each other's activities). In a close perspective, Hutchins and Klausen (1992) develop the idea that cognition is fundamentally distributed. The use of artifacts is analysed to identify their properties highlighting the mental operations they support. So collective cognition and artifacts are not considered independantly.

Considering such studies we suggest that collective cognition has to be seen in its interaction with environmental resources used by human agents. Cooperative activities imply that agents communicate in order to share their understanding and to recognize their intentions in face-to-face situations. The problem we address in this paper is to assess the parts of non-verbal and verbal communication in cooperation. The idea that non-verbal aspects in communication are informative seems to be a trivial one, but identifying their role becomes an important question, especially if we are concerned by interface design. In studying interaction between experienced human agents, in a face-to-face situation, our hypothesis is that non-verbal resources are needed in addition to verbal ones to ensure communication, for the following reasons.

Communicating partners are faced with the difficulty of constructing and updating a common cognitive environment which enables them to cooperate. As shown by Sperber and Wilson (1989), communication is based on inferential and decoding processes, so the success of communication is uncertain, and the construction of a compatible meaning of the situation implies that human agents continuously regulate or anticipate misunderstandings which can arise in interpreting the other's utterances and actions. The part played by non-verbal resources becomes essential in such time-constrained situations where verbal interventions, needed in particular to recover a failure in mutual understanding, appear for several reasons to be only partially adapted.

So our purpose is to examine the ways in which modalities of verbal and well as non-verbal communication are required in order to establish a shared understanding in symmetrical situations of interaction (both are experienced controllers). Firstly, our aim is to discuss the model of communication underlying the study of cooperation. Secondly, in this perspective, we will consider communications in their verbal as well as non-verbal modalities.

## 2 Mutual cognitive environment

Communication is a co-construction process where achievement is not certain. In order to describe the process of building a shared understanding, two main models of communication can be mentioned. On the one hand, with the Shannon and Weather studies (1948), communication is considered as a coding-decoding process. On the other hand, Sperber and Wilson

(1989) question the mechanical nature of this process and highlight the inferential nature of communication. From the first point of view, establishing common knowledge or beliefs would imply that each agent shares the same knowledge used to code and decode a message. The problem in this case is to determine the level of mutual knowledge required (agent A knows that agent B knows that agent A knows that agent B knows...).

An alternative model proposed by Sperber and Wilson (1989) allows a more specific description of what shared information actually is, using the concepts of cognitive environment and the relevance principle. The cognitive environment is an individual construction elaborated by each agent through information acquired in his/her environment according to his/her beliefs, personal theories, etc. and to his/her perceptive and inferential abilities. The mutual cognitive environment is based on the hypothesis formed by the agents. This model presents communication as a process based on an imperfect heuristic. The agents are not able to determine accurately the respective cognitive environment. Moreover, this model takes into account ostensive behaviour. Environment resources may support intention recognition and modify cognitive environment. The heuristic nature of communication implies that the main problem faced by human agents is to ensure mutual understanding. A major problem is that agents have to manage numerous misunderstandings which can arise in the co-construction of a compatible meaning of the situation. Several aspects have been studied in this perspective. Rogers (1992) shows that the informal mechanisms of failure of coordination have to be examined to develop resources which can facilitate the detection of misunderstanding. In the case of a networked environment, engineers need to know what others are doing on the network in order to manage communicative problems which are very time-consuming. In face-to-face situations, several communicative resources can be used by participants to construct a compatible meaning of the situation.

I. The establishment of a mutual cognitive environment using verbal resources

Assumptions about the other's knowledge and beliefs on the basis of verbal resources are needed to communicate. The establishment of shared knowledge is analysed by Clark and Wilkes-Gibbs (1990) in verbal tasks (tangram). Experiments show that to identify the reference of a given expression, the subjects base themselves on the representation they have of the other's knowledge, their performance evolves in the course of a session. Krauss and Fussel (1990) point out how agents try to determine what is mutually shared in order to communicate. Krauss and Fussel (1990) evoke three interrelated sets of mechanisms which communicators use to establish the existence of common ground: direction knowledge (assumption that the partner is able to have the appropriate interpretative context, because of co-presence, for example); category membership (such as prediction about individual knowledge in respect to his/her social category); and interactional dynamics (for example, what has been! said is assumed to be known). The dynamics of interaction and, in particular, the part played by feed-back has been developed by Clark and Schaeffer (1989), Krauss and Fussel (1990), Clark and Brennan (1991). Clark and Brennan (1991) analyse the contributions of agents in conversation divided into two steps, a presentation and an acceptance phase. For Clark and Brennan (1991) all collective actions, and in particular communication, are based on assumption of shared ground (mutual knowledge, beliefs and assumptions) which is constantly being updated. The grounding process evoked, (a collective process by which the participants try to reach a mutual belief) takes particular shape with respect to face-to-face situations.

So a major problem occurs in time-constrained situations, this is how verbal resources will be used when verbal explanations, reformulations, etc. (ensuring mutual understanding) become costly.

II. Establishment of a mutual cognitive environment using non-verbal resources

In the communication process, meaning emerges from the interpretation of verbal utterences, but also from non-verbal elements which contribute to modification of the cognitive environment of human agents. As Shapiro et al. (1989), Hughes et al. (1992) have shown, the strip is essential in the social organization of work in the team, in which activities and information are distributed and used among members. Strips are updated according to the usual routes, symbols, circles around relevant destinations (e..g. climbing or descending aircraft, etc.) They compose an evolving history and a plan of the controller's intentions and decisions. On a wider scale, through studies of interaction between human agents co-present at the work station, it appears that external resources available at the work station provide a support for collective cognition. Heath et al. (1993) show that the ways in which dealers coordinate their actions and participate in each other's conduct (in the dealing room of a City of London international securities house), is linked to the co-presence which allows the operators to collaborate. Initiation of mutual engagement, for example, is based on the direction of the operators' looks and their body postures. Other studies have shown how resources in the working environment are used to determine, for example, the availability of colleagues, which condition the dialogues between permanent staff and doctors in the SAMU emergency services (Benchekroun et al., 1993). Studying the grounding process in communication, Clark and Brennan (1991) show in the case of establishing the referential identity (mutual belief that the addressees have correctly identified the referent) that several techniques are used, one is indicative gestures (pointing, looking or touching). As indicated by Krauss and Fussel (1990) literature provides little support for affirming that communication, when conveyed by visual and verbal channels, is more efficient than by verbal channels only (Krauss and Fussel, 1990, p.138). However, their study in using refential communication tasks shows that inserting a delay and temporally displacing feedback response is enough to demonstrate the extent of the commmunicator's dependance on feedback to formulate efficient referencing expressions. Other experiments show that visible feedback, (smiles, head shakes and nods) can compensate for the absence of verbal information (so with visible information available, the effect of delay! ed transmission is decreased). This introduces the part that could be played by non-verbal aspects in the success of communication. In this perspective, artefacts considered as environmental resources can be used in order to communicate. A major problem occurs, how will non-verbal resources be used to ensure successful exchanges? Other questions also arise: can non-verbal resources be trusted in the sense that they are reliable for mutual understanding? What level of mutual understanding could be reached by non-verbal communication?

## 3 Multimodality and communication

As an introduction, two dimensions of non-verbal resources will be discussed, firstly the status of non-verbal resources (contextual dimension versus specific modality) and secondly the informative nature of non-verbal resources (versus communicative).

I. A first presupposition could be to consider the non-verbal dimension of interaction as part of the context of interpretation of the verbal communiction or on the contrary, as a specific mode of expression.

A first attempt to understand the part played by non-verbal resources could be in identifying how they are used as a context in interpreting verbal utterances. From another angle, non-verbal resources can be analysed with respect to the specific meaning they convey. In non-verbal communications studies, it appears that non-verbal resources cover several dimensions. The "multi-channel" notion of human communication by Cosnier and Brossard (1984) characterizes face-to-face interactive situations, giving the idea that the agents transmit a total heterogeneous message, resulting from a combination of several elements (voco-acoustic, visual and olfactory, tactile and thermic). These authors (1984) establish a distinction between two mimetic-gestural functions (1) framing the interaction, the mimetic-gestural function is assimilated as indication of the context (in that it provides a situational context); (2) as "co-text" it then makes a dynamic contribution to the exchanges. In t! he same way, the notion of configu rations of multi-channel signs is developed by Scherer (1984). In a similar way, Cadoz (1993) shows that some semiotic body expressions (e.g. informative messages destined to the environment) are actions combined with verbal expression and can be considered as part of the communicative structure and a specific mode of expression.

II. A second reading of non-verbal messages belongs to the informative or communicative dimension.

The elaboration of collective cognition in working situations goes through various modalities of communication; for example, an action (pointing to the radar scope) may be interpreted by co-workers as an act of communication, aiming at this emphasis of relevant information. This action can be analysed from several points of view. On the one hand, the action is a means for each agent to organize information for him/herself. On the other hand, the action may be performed in order to communicate (for example, showing something to a partner and saying "have you seen this?"; the deictic gesture is part of communication). A more extensive analysis of the intentional nature of behaviour in human communication leads us to distinguish two intentional levels described by Sperber and Wilson (1989). The first informative intention appears when a speaker makes elements of a situation manifest without showing the intentionality of making them manifest. For example, an agent, rather than a! sking his colleague for some help to repair his tools, displays the various components prominently. The second communicative intention corresponds to an interactive situation in which the speaker informs his interloctur and then manifestly shows that he has the intention of requesting something.

In face-to-face interactions between controllers, several communictive modalities can be employed according to the nature of the information communicated (diagnostic of a conflictual situation, transfer conditions of an aircraft to an adjacent sector, etc..) and the status of the information (urgency of the situation, work context, etc.) At the work station, the information is processed and memorized through the various supports, for example, the organization and handling of strips. Certain properties of the strips have been demonstrated elsewhere, see Shapiro et al. (1989), for example its mobility and writability, and the visual accessibility of this information. In the same way, the agents can listen to radiotelephony communications to be aware of each other's activities. These elements will be considered through their implications on communication in the team.

First, the analysis of the exchanges between controllers reveals that verbal and non-verbal resources contribute to the establishing and updating of a mutual cognitive environment related to the current situation in various ways: non-verbal actions are used by the controllers to provide pieces of information which they do not communicate verbally (for example, urgency not verbally expressed or the figurative aspect of a particular regulation between aircraft).

Secondly, with respect to the heuristic nature of human communication (see Sperber and Wilson, 1989), multimodality appears as a resource which is used by human agents to anticipate or regulate misunderstandings in such face-to-face situations (for example, strip moving combined in exchanges with the name of the flight concerned, which is not verbally expressed). The complementarity of verbal and non-verbal resources seems essential for ensuring mutual understanding (for example, non-verbal actions are used by the controllers to provide a context which their partner will use to interpret the meaning of what is verbally said).

Third, our present study emphasizes that communications between human agents are supported by the use of verbal and non-verbal resources for establishing a mutual cognitive environment at an informative level and at a communicative level. The actions and utterances of human agents, produced when carrying out their own activity, are potentially used at the same moment by other agents for intention perception (for example, the executive controller writing the present flight level on the strip is seen by the planning controller).

## 4 Study situation

The team studied is an operational unit, composed of the executive controller and the planning controller in a face-to-face interactive situation, with the controllers of adjacent sectors in telephone contact with the planning controller and the various pilots in radiocontact with the executive controller. The executive controller's responsability is sectoral, including the maintenance of separation standards within the sector. The planning controller is in charge of coordination of the traffic passing in and out of the sector and acts on the stream of traffic received into the sector. The controllers organize the flow of traffic to avoid conflicts between flights, including constraints like the need for flights to be expedited as soon as possible (fuel consumption) and some constraints linked to the situation, such as meteorology. The training of both controllers is the same, including the theoretical aspects, (for example, procedures described in Air Traffic Services Manuals) concerning each sector and finally training in real situations under the responsability of an experienced controller. Air traffic regulation is a complex task because it implies decision-making and resources management with time constraints, the processing of large quantities of information, which are both evolving and uncertain (Leroux,1992), functional and temporal coordination between actions performed by each agent, etc. For experienced controllers, the problem, except for defining a strategy to solve a conflictual interaction, is to envisage the consequences of this solution on the surrounding traffic (to avoid creating other conflicts) and to monitor the application (acting at the right moment, checking that the aircraft did actually turn). The monitoring of traffic is a high cost for the controller (Leroux,1992), ("letting a situation evolve") one of the risks being not acting early enough on a conflictual interaction in a context where attention is shared between several conflictual interactions at the same time. Complexity arises from the dynamic nature of the environment, each dec! ision has to be evaluated in respe ct of the evolving state of the traffic, requests from pilots and adjacent sectors, and unexpected events have to be taken into account as soon as possible. Each agent has to take decisions under time pressure, some of them (negotiation with controllers from other sectors, for example) have consequences on the actions of others, in this perspective communications are needed but there is also the need to wait until the colleague is available to inform or negociate with him/her. Controllers use a number of information sources: the flight progress board (including strips arranged in front of the executive controller himself, according to problems detected

# DECLARATION OF TED BALDWIN CONCERNING THE "PAAM 96: PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON THE PRACTICAL APPLICATION OF INTELLIGENT AGENTS AND MULTI-AGENT TECHNOLOGY, 22ND-24TH APRIL 1996" REFERENCE

I, TED BALDWIN, declare as follows:

1. I am over the age of 18, have never been convicted of a felony or crime of moral turpitude and am legally competent to make this declaration. I have personal knowledge of the matters stated herein.

2. I am a librarian at the University of Cincinnati Libraries ("Library") located within the University of Cincinnati in Cincinnati, Ohio.

3. I have been employed by the Library for 18 years.

4. I am familiar with the regularly conducted activities of the library, including the record-keeping practices of the Library, as well as the indexing, cataloging and circulation of the books and references in the Library's collection.

5. In the late 1990's and early 2000's, the Library regularly maintained physical records in the course of its operation. One such regularly-maintained record was a catalog entry corresponding to items in the Library's collection. These entries are maintained in the Library's catalog. Another such regularly-maintained record was an intake stamp, which indicated the date a new item was indexed in the Library's catalog and made available to the public in the Library's stacks. The intake stamp was typically affixed on the inside of the item at or near the date the item became publicly available.

6. The regularly-maintained records of the Library demonstrate that PAAM 96: Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, 22nd-24th April 1996, including the article. "Development Tools for the Open Agent Architecture" by D. Martin, A. Cheyer, and G. Lee on pages 387-404, was placed in the Library's collection in accordance with the Library's regular practices and made publicly available on or about November 17, 1997.

7. An accurate copy of the reference discussed in paragraph 6 is attached as Exhibit A. The copy accurately depicts the intake stamp affixed to the original, showing that the intake process discussed above in paragraph 5 was completed for this reference on or about November 17, 1997.

8. The Library's catalog referenced in paragraph 5 above is available for members of the public to search. The catalog can be searched by subject, title, author, and keywords. An entry for the reference discussed in paragraph 6 has been maintained in the Library's catalog since on or about November 17, 1997. An accurate copy of this catalog entry is attached as Exhibit B.

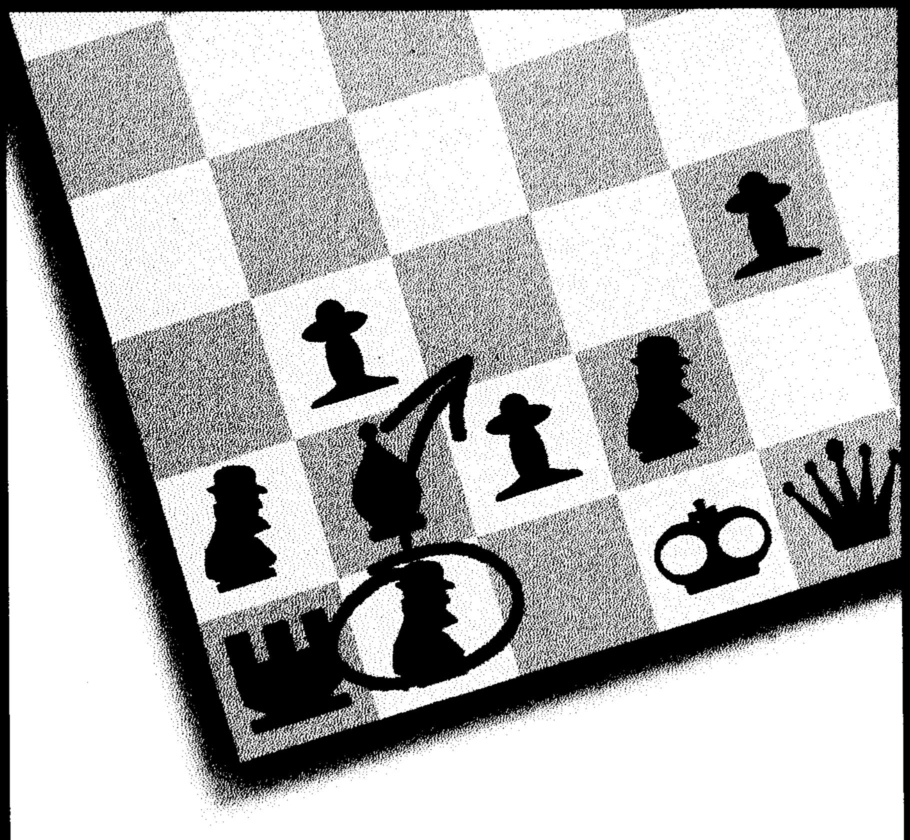
All statements made herein of my own knowledge are true and all statements made on information and belief are believed to be true. I further understand that willful false statements and the like are punishable by fine or imprisonment, or both under Section 1001 of Title 18 of the United States Code. I declare under penalty of perjury of the laws of the United States that the foregoing is true and correct.

Executed on Nov 29, 2017 in Cincinnati, OH.
[DATE] [CITY/STATE]

Ted Baldwin

EX. A

# PAAM 96



The First International Conference and Exhibition on

# The Practical Application of Intelligent Agents and Multi-Agent Technology

The Westminster Central Hall, London, UK
22nd - 24th April 1996

## PROCEEDINGS

PAAM 96

# Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology

**Conference Organisation**

The Practical Application Company

**Programme and Tutorial Chairs**

Barry Crabtree and Nick Jennings

**Sponsorship Co-ordinator**

Clive Spenser

22nd-24th April 1996

Westminster Central Hall, London , UK

PAP, P.O. Box 137, Blackpool, Lancashire, FY2 9UN, UK
Tel: 44 (0)1253 358081, Fax: +44 (0)1253 353811
Email: proceedings@pap.com

ISBN 0 9525554 3 3
Published by The Practical Application Company Ltd

# Development Tools for the Open Agent Architecture*

David L. Martin, Adam Cheyer
SRI International

Gowang-Lo Lee
ETRI

### Abstract

The agent-based paradigm for software systems cannot realize its full potential, and will not become widespread, until adequate agent development tools and environments are available. To address this need, an exploration of the requirements for such tools and environments has been conducted in the context of the Open Agent Architecture (OAA) project, and has resulted in the creation of the Agent Development Toolkit (ADT). The ADT provides a variety of mechanisms that support the specification and implementation of individual agents, as well as cooperating communities of agents. Special attention has been given to tools that enable an agent developer to construct intelligent user interfaces, which allow users to express their requests of agents using spoken and written natural language in combination with other modalities. This paper discusses a number of general requirements that were identified for agent development environments, reports on the design and functionality of the ADT, and shows how the ADT addresses those requirements. In addition, we describe our experience to date in constructing OAA-based agent systems, and future directions in extending the ADT.

## 1 Introduction

A number of important and interesting investigations have recently been made into the languages, architectures, algorithms, and formal analyses of agent-based systems, and substantial agent-based systems are being fielded in a variety of domains. There are good reasons for this. The notion of autonomous, cooperative, and intelligent agents as fundamental system building blocks provides an evocative metaphor and a natural paradigm for harnessing explosive increases in interconnectivity and information access. From a system developer's perspective, this paradigm holds the promise of constructing flexible, adaptable systems that provide intelligent services based on the cooperative

---

*This paper was supported by a contract from the Electronics and Telecommunications Research Institute (Korea). The first author can be reached by email at martin@ai.sri.com.

efforts of the most capable and most appropriate agents for the job at hand, selected from a potentially vast array of distributed software and hardware resources. [1]

While the results of these investigations provide many valuable elements of infrastructure for agent-based systems, it must be recognized that the agent-based approach cannot realize its full potential, and will not become widespread, until adequate agent development tools and environments are available. To date, very little has been done to address this need.

There are a number of interesting questions to be addressed: What new requirements and challenges arise for development tools that are unique to agent-based systems? How does the inherent autonomy and loose coupling of agents affect the development process and the resulting artifacts such as documentation? How can we best facilitate the construction of a collection of interoperable agents written in various languages and operating on various platforms, and agents derived from existing applications and legacy information sources? How much of the creation of an agent-based system can be automated?

An agent system that provides an intelligent user interface — allowing users to express their requests by using spoken and written natural language in combination with other modalities — raises additional challenges regarding development environments. For example, one important question is how best to provide support for the agent developer, who is not likely to be a computational linguist, in tailoring the linguistic processing components of the system to handle the domain-specific expressions that may be expected to appear in users' requests. [2]

An exploration of these questions has been conducted in the context of the Open Agent Architecture (OAA) project, and has resulted in the creation of the Agent Development Toolkit (ADT). This paper is concerned with the requirements that motivated the creation of the ADT, and the functionality that evolved to meet those requirements. The following section presents a general discussion of requirements that are characteristic of development environments for agent-based systems. In Section 3 we give an overview of the OAA, and of results to date in constructing OAA-based agent systems. Section 4 shows in some detail how many of the requirements mentioned earlier have been addressed by the ADT. Finally, in Section 5, we draw conclusions and mention some current and proposed work to extend the OAA and the ADT.

## 2 Challenges for Agent Development Environments

In highlighting some of the general requirements and challenges that can be identified for development environments for agent-based systems, we are not attempting to give

---

[1] Because of the wide variety of systems to which the word 'agent' has been applied recently, it may be helpful to indicate what we mean by 'agent-based system'. The type of system we have in mind is one in which the services provided are accomplished through the cooperative efforts of a number of independent software processes, each of which is persistent and acts with a high degree of autonomy.

[2] Most other important areas of exploration in agent-based systems — learning, mobility, negotiation, and so forth — also introduce new challenges for development environments.

an exhaustive list. We do believe that the points mentioned here are applicable to most agent-based systems. In describing the Open Agent Architecture in Section 3, we will be able to show in greater detail how these requirements arise in that particular context, and in Section 4 we show how they are addressed by the ADT.

## 2.1 Supporting Conformance

Because of the emphasis on interoperability inherent in agent-based systems, there is a critical need for each agent to be designed so as to interact correctly (that is, in accordance with protocol) with the other agents in the system. Thus, an agent development environment should guide the developer in adhering to the protocols used by the system.

Some form of this requirement has existed in all software development paradigms; after all, even in the simplest programs, procedure calls must match the appropriate procedure declarations. However, the need for conformance is likely to be more strenuous in agent-based systems, in two respects. First, agent programming interfaces and interactions between agents — and hence, the protocols for specifying these — tend to be more complex than interfaces and interactions between the elements of systems built using traditional approaches. Second, it is a goal of most agent systems that the development teams of the various agents be able to work independently, remotely, and on widely heterogeneous platforms — but while incurring as little overhead as possible due to the interdependencies of agents.

This requirement of conformance applies as strongly to agent *documentation* as it does to agent coding. In particular, the ongoing evolution of an agent-based system by widely distributed and independent groups of developers will require documentation of available agents and their capabilities in a consistent, automatically searchable format.

## 2.2 Supporting Heterogeneity

In agent development, as in most software development, conformance and heterogeneity are two sides of the same coin: it is precisely because of the need to achieve a meaningful level of interoperability between widely heterogeneous agents that it is critical for agents to conform to the same protocols.

Many different types of heterogeneity can occur in an agent-based system. Three that are of concern from the agent developer's point of view are the multiplicity of implementation languages, the multiplicity of execution platforms, and the mixture of newly created agents with those that have been adapted from legacy applications or information sources.

Thus, the design of an agent development environment (as well as the design of the architecture) should allow for an equal level of support for an agent's development, regardless of its language, platform, or origin.

## 2.3 Construction of Agent Communities

An agent-based approach encompasses a new definition of "system" (or at least a definition modified in some important ways), and consequently calls for new conceptualizations of what it is to create a "system". Agent-based system construction involves the identification of a set of agents that can do a job together. Wherever possible, parts of a system's functionality are provided by reuse of existing agents, but in any case the determination of what services are provided by existing agents is an essential prerequisite to the design of new agents. Thus, a development environment should make it as easy as possible to manipulate (e.g., locate, browse, inspect, visualize) agents as the basic building blocks of systems. In particular, it should provide support for identifying the capabilities of existing agents. It should also provide support for specifying new configurations of agents for interoperation.

## 2.4 Running and Debugging Systems

Agent-based approaches also entail changes in what is meant by "system execution". Invoking — and monitoring — an agent-based system can become much more involved than it is under today's predominant software paradigms. Rather than focusing on the behavior of a single process, or a tightly regimented series of client-server interactions, the agent-based system developer needs to be able to initiate and ensure the continued availability of an entire collection of processes running in diverse environments. He must be able to view the global activity of the collection, as well as the local activities of specific agents. These needs call for more powerful execution and debugging aids than currently exist. Thus, an agent-based development environment should provide new mechanisms for instantiating, monitoring, and debugging operational configurations of agents. Agent-based debugging aids will most likely be constructed on models borrowed from the field of simulation.

## 2.5 Facilitating Use of Support Agents

In our terminology, a *support agent* is one that provides services of great importance to many, if not most, agents operating in a system. Thus, while not a fixed part of the agent system infrastructure, a support agent is thought of as having a more fundamental status than an ordinary application agent, because of the widespread demand for its use. Because of the emphasis in the OAA on intelligent user interfaces, speech recognition and natural language understanding agents have become two very important examples of support agents in the OAA.

Support agents pose special problems for agent development tasks because in many cases they employ sophisticated techniques. As a result, customizing a support agent for a particular task domain is likely to require substantial expertise — a level of expertise that the average agent developer may not possess and may not have the time to acquire.

Because of their quasi-standardized use with the system, however, support agents offer an opportunity to provide knowledge-acquisition tools that support their use. For example, as we show in Section 4.2, the use of speech recognition and natural language understanding agents can be supported with tools for the introduction of natural language vocabulary and concepts relevant to each agent that employs their services.

## 3 The Open Agent Architecture

The Open Agent Architecture provides a framework for integrating a society of software agents, each possessing a high degree of independence and autonomy, within a distributed environment. A collection of agents satsifies requests from users, or other agents, by acting cooperatively, under the direction of one or more facilitators (which are themselves agents of a special type).

The system's architecture, based loosely on Schwartz's FLiPSiDE system [7], uses a hierarchical configuration in which each application agent connects as a client of a facilitator. Facilitators provide content-based message routing, global data management, and process coordination for their set of connected agents. Facilitators can, in turn, be connected as clients of other facilitators. Each facilitator records the published capabilities of their subagents, and when requests arrive (expressed in the Interagent Communication Language, described below), the facilitator is responsible for breaking them down and for distributing subrequests to the appropriate agents. An agent satisfying a request may require supporting information, and the OAA provides numerous means of requesting data from other agents or from the user.

Agents share a common communication language and a number of basic structural characteristics and capabilities. An agent library provides this common functionality. For example, every agent can install local or remote triggers on data, events or messages; manipulate global data stored by facilitators; and request solutions for a set of goals, to be satisfied under a variety of different control strategies. In addition, the agent library provides functionality for parsing and translating expressions in the Interagent Communication Language, and for managing network communication using TCP/IP. Agents may be implemented (or derived from existing applications) in any programming language to which the agent library has been ported, and may run on any network-linked platform.

The OAA has been described in greater detail in [4].

### 3.1 The Interagent Communication Language

The OAA's Interagent Communication Language (ICL) is the interface language shared by all agents, no matter what machine they are running on or what computer language they are programmed in. The ICL has been designed as an extension of the Prolog programming language, in order to take advantage of the power of unification and backtracking during interactions among agents.

Every agent participating in an OAA-based system defines and publishes a set of capabilities specifications, expressed in the ICL, describing the services that it provides. These establish a high-level interface to the agent, which is used by a facilitator in communicating with the agent, and, most important, in delegating service requests (or parts of requests) to the agent. Partly due to our use of Prolog as the basis of the ICL, we refer to these capabilities specifications as *solvables.*

For example, in creating an agent for a mail system, solvables might be defined for sending a message to a person, testing whether a message about a particular subject has arrived in the mail queue, or displaying a particular message onscreen. For a database wrapper agent, one might define a distinct solvable corresponding to each of the relations present in the database.

## 3.2 StartIt

As mentioned in Section 2.4, agent-based architectures introduce strenuous requirements for invoking and monitoring systems of agents. StartIt addresses these requirements, and provides an important bridge between the functionality of the ADT and that of the OAA.

Once a collection of interoperable agents has been assembled to work on a set of tasks, StartIt provides the means of invoking each of the agents on the correct platform, according to the system protocols of that platform, and ensuring that the agent makes the required connection to an OAA facilitator. Of equal importance, StartIt monitors the status of each agent to see that it continues to function correctly. In the event that StartIt detects a failure of one of the agents, it is able to take steps to recover from the failure and automatically restart the agent.

Startup specifications for each agent and instructions on how to deal with failures are contained in configuration files which, as described below, can be automatically generated by a component of the ADT.

## 3.3 OAA-Based Prototype and Fielded Systems

The OAA has been used as the framework for a number of applications in several domain areas. The first OAA-based system was a multifunctional "office assistant", in which fourteen autonomous agents provide monitoring, communication and management capabilities for business applications such as online calendars, electronic mail, or databases [4]. In a typical scenario, agents with expertise in email processing, text-to-speech translation, notification planning, calendar and database access and telephone control cooperate to find a user and alert him or her of some important message.

The OAA has also been used to construct flexible and natural user interfaces to agent-based and conventional applications. In the CommandTalk system, currently installed at the Marine Corps Air Ground Combat Center at Twentynine Palms, CA, a collection of OAA-enabled agents provide a spoken-English interface to a map-based simulation of armed forces. Another OAA-based multimodal user interface project focuses on techniques for merging simultaneous streams of pen and voice input to form multimedia queries about data retrieved from commercial Internet web sites [2].

## 4 The Agent Development Toolkit

The Agent Development Toolkit, or ADT, is built around three loosely coupled core components, and presents itself via a user interface component.

- The Programmer's Agent Construction Tool (ProACT) is used by an agent designer to define and maintain the capabilities and other properties of an agent, to manage documentation for the agent, and to generate a code template for the agent.
- The Linguistic Expertise Acquisition Program (LEAP) facilitates the task of interfacing a new agent with existing linguistic support agents such as natural language parsers and speech recognition systems. This involves obtaining semantic information about the domain in which the agent operates, the services provided by the agent, and the English words that will be useful in composing requests for these services. To make these words useful to the system, LEAP extracts from the agent developer information about their linguistic attributes; it does so by asking the developer simple questions about how and when those words are used. Once the linguistic knowledge has been acquired, LEAP generates or updates the appropriate knowledge bases needed by the linguistic support agents.
- PROJECT allows the developer to create and maintain repositories of reusable agents, and to choose from available repositories to produce an operable configuration of agents for a particular application domain. Once the configuration has been selected, PROJECT can produce a configuration file for use by StartIt, the OAA's system execution manager.
- The user interface component provides integrated access to the features of all three core components. It provides editing capabilities for the artifacts of each core component, such as agent specifications, iconic representations of agents, source code, domain classes and vocabulary, agent repositories, and project configurations.

The ADT has itself been constructed within the OAA. That is, each of the three core components, as well as the user interface, is instantiated as one or more OAA agents. Thus, in constructing the ADT, we were able to take advantage of the benefits of the agent-based paradigm. For example, we were readily able to use a mixture of languages and platforms (some under UNIX [3] and some under Microsoft Windows) in implementating the components. In particular, the user interface benefited from the use of rapid development user interface tools available under Microsoft Windows, and LEAP benefited from being implemented under UNIX, where we were able to make good use of our Prolog

---

[3] All product names mentioned in this document are the trademarks of their respective holders.

development environment and some existing source code from related projects. Further, the use of the OAA ensures future extensibility via the addition of new agents.

In the following discussions of the three core components, the use and appearance of the user interface component is not covered in detail, but parts of it are mentioned in the core component descriptions, and parts are shown in the accompanying figures.

## 4.1 ProACT: Defining and Constructing Agents

ProACT guides an agent developer through the various phases of agent creation and maintenance.

An agent developer starts creating a *new* agent by defining, in ProACT, its name, author, title, version number, and icon. To inspect or modify an *existing* agent, the agent can be opened using either of two familiar techiques: the existing agent's specification file can be selected from a file navigation dialog, or its icon can be selected from those in the currently selected agent repository. (Agent repositories are selectable using PROJECT.)

The agent programmer can then use ProACT to enumerate the agent's capabilities in terms of the Interagent Communication Language. The ICL editing window provides an opportunity to ensure conformance to protocol, by performing syntax checks and prompting the developer for missing syntactic elements. [4]

Once the capabilities of the agent have been specified, ProACT encourages the agent programmer to provide documentation for the agent, in a standardized format. Information may be entered using built-in documentation editors, which provide templates for describing the agent itself, and each of the agent's capabilities specifications. After documentation has been edited, ProACT automatically generates HTML representations of the information that can be published on the World Wide Web, and thus can be made readily available to other agent developers collaborating on the project, or those who may add agents to the project at some future time.

The use of HTML as a documentation medium is motivated by the requirement, discussed earlier, to support widely distributed teams of agent developers with up-to-date specifications that can automatically be searched for reusable agents providing some needed service. Publishing documentation in HTML allows developers to employ any of a wide variety of available Web tools. For example, ProACT interfaces with Harvest [1], an Internet tool for indexing and searching Web pages. In the Harvest framework, brokers and gatherers can be set up to collect all published OAA documentation from anywhere in the world, or from selected subgroups of agent development sites — thus providing an efficient query mechanism to search for appropriate agents for reuse.

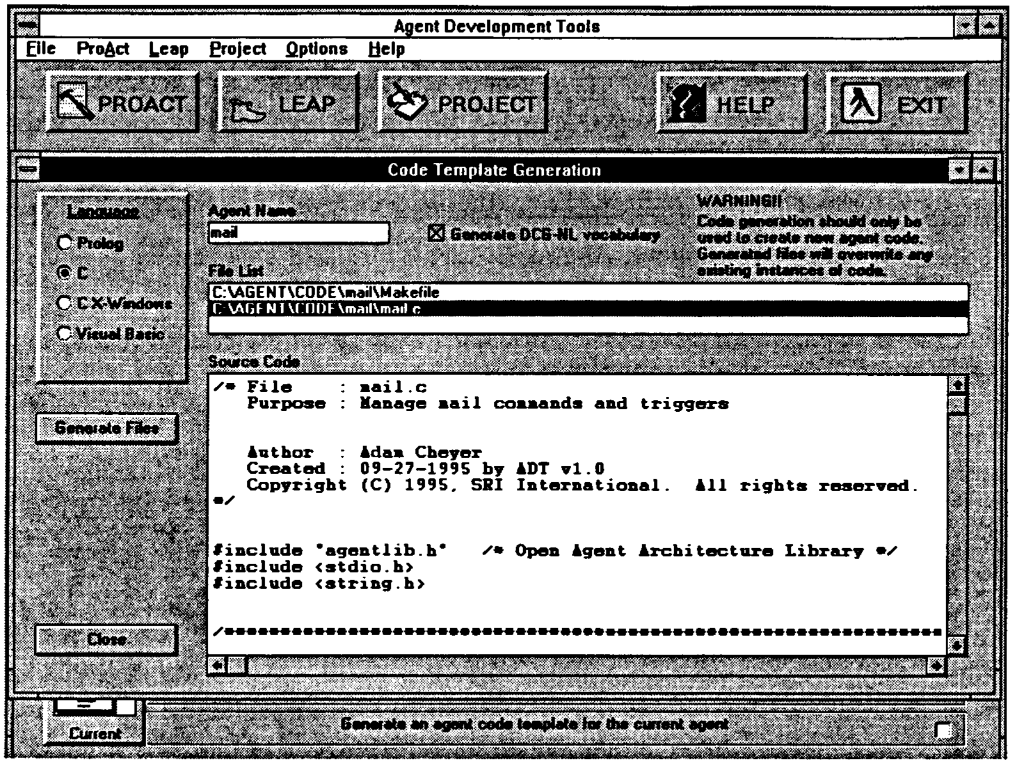
ProACT supports heterogeneity by generating code templates for agents in several programming languages, currently Prolog, C, C with X Windows, and Visual Basic. Delphi and Lisp will be added soon, as libraries in these languages have recently been added to the OAA. Code template generation is a useful function for the novice programmer, who

---

[4] As of this writing, these syntax checks are under development.

may not know all the intricacies of building a new agent, as well as being a timesaver for the expert user. Code template generation is also convenient when an existing agent is ported to another programming language.

A ProACT screen is shown in Figure 1. In this figure, code template generation, in C, has just been completed for a new agent.



○ Prolog
● C
○ C X-Windows
○ Visual Basic

Agent Name: mail

☒ Generate DCG-NL vocabulary

WARNING!! Code generation should only be used to create new agent code. Generated files will overwrite any existing instances of code.

File List:

C:\AGENT\CODE\mail\Makefile
C:\AGENT\CODE\mail\mail.c

Source Code:

```
/* File      : mail.c
   Purpose : Manage mail commands and triggers


   Author   : Adam Cheyer
   Created : 09-27-1995 by ADT v1.0
   Copyright (C) 1995, SRI International.  All rights reserved.
*/


#include "agentlib.h"   /* Open Agent Architecture Library */
#include <stdio.h>
#include <string.h>


/*****************************************************************
```

Generate an agent code template for the current agent

Figure 1: Using ProACT to generate source code for an agent.

## 4.2 LEAP: Adding Speech and Natural Language Understanding to Agents

Agents provide functionality that can be accessed by other agents, by the user through a graphical user interface, or sometimes by the user through a natural language (spoken or written) interface. As mentioned in Section 2.5, speech recognition and natural language processing capabilities are made available to all agents in the OAA by specialized support agents.

To provide a natural language interface to an agent, the agent designer must generate linguistic knowledge bases for the Natural Language and Speech Recognition agents, which enables these agents to handle spoken and written requests that are appropriate for the agent. LEAP is a tool for guiding the user through this process, and is primarily concerned with the requirements expressed in Sections 2.1 and 2.5.

It is important to realize that the roles of the Speech Recognition and Natural Language agents can be played by different agents in different OAA configurations (indeed, it is possible to have several different Speech Recognition and/or Natural Language agents operating within a single configuration). These Speech Recognition and Natural Language agents can be of varying levels of sophistication, and in some configurations, there are advantages to using relatively simple approaches (for example, some configurations have employed Natural Language agents based on Prolog Definite Clause Grammars). However, in most settings, one wants to use the most powerful, flexible approaches available, and thus our efforts have been focused on the use of two very sophisticated systems developed at SRI: the Decipher [3] speech recognition system, and the Gemini [5] natural language understanding system, both of which have been used as agents in a number of OAA-based systems. Consequently, the requirements for LEAP have largely been driven by these two systems.

Although the Speech Recognition and Natural Language agents provide considerable flexibility in specifying knowledge for new domains, they were written by and for computational linguists. Consequently, extending the domain knowledge and linguistic knowledge of these support agents (as is true of most powerful speech recognition and natural language systems) has heretofore been a complex task requiring expertise in computational linguistics. This has been an acceptable requirement in their original context of use. However, their use within the OAA creates a new context, characterized by the following conditions:

- New and widely varying domains are added frequently.
- As agents are introduced and developed in a domain, the knowledge needed by the Speech Recognition and Natural Language agents changes rapidly and may continue to evolve over a long period. This change involves knowledge of linguistic usage as well as knowledge of the solvables (agent capabilities descriptions) currently made available in the domain.
- Agent developers, rather than linguists, will introduce new domain knowledge to the Speech Recognition and Natural Language agents.

LEAP's goal, then, is to assist the nonlinguist in introducing new domain and linguistic knowledge to Speech Recognition and Natural Language agents.

### 4.2.1 LEAP's Subcomponents and General Approach

LEAP's mission involves acquiring four types of knowledge: domain knowledge, as captured in a class hierarchy; knowledge of the solvables provided by the agents being used in an OAA-based system; some types of linguistic information (morphological, syntactic, and semantic) about the vocabulary that may be used in formulating requests of the agents; and phonetic (pronunciation) information about this vocabulary.

The first three of these knowledge types provide the critical connections that the Natural Language agent will need (at execution time, not at agent development time) to transform an English request into a formal goal that may be handled by an OAA facilitator. This goal, an expression in a first-order logical notation, contains solvables as subgoals. The facilitator, in satisfying the goal, will dispatch each solvable to an agent that can handle it. The fourth type of information will be used (also at execution time) by the Speech Recognition agent in recognizing spoken requests.

LEAP has a subcomponent corresponding to each of these four types of knowledge; these subcomponents are the Class Hierarchy Editor, the ICL-NL Linker [5], the Word Wizard, and the Pronunciation Wizard.

The sequence of events for telling LEAP about a new agent is as follows: First, using the Class Hierarchy Editor, inspect and edit the class hierarchy to ensure that the types of objects the agent deals with are represented in the hierarchy. Then, using the ICL-NL Linker, provide semantic information about the agent's solvables (these have already been entered, using ProACT). Next, using the Word Wizard, enter words that are expected to be contained in users' requests for the agent. Finally, for any words for which the Pronunciation Wizard doesn't already have a phonetic description, use the Pronunciation Wizard to select and/or edit one.

In our presentation, here, of the first three subcomponents of LEAP, we are primarily concerned with operations that help to satisfy the knowledge base requirements of the Natural Language agent. This is because its knowledge base is considerably more complex than that required by the Speech Recognition agent. Indeed, most of the information required by the Speech Recognition agent can be viewed as a subset of that needed by the Natural Language agent. One notable exception to this, however, is the information gathered by the fourth subcomponent, the Pronunciation Wizard.

### 4.2.2 LEAP's Class Hierarchy Editor

Nearly all rules in the knowledge base of the Natural Language agent refer to the classes defined in the class hierarchy. The class hierarchy is a tree that contains what the Natural Language agent recognizes as the primitive conceptual categories to which entities may belong, and expresses the superclass and subclass relationships that hold between them. Higher levels of the hierarchy contain the more domain-independent classes, whereas lower levels tend to be more domain-specific. For example, the class *agent* — a class likely to be near the root of the hierarchy — might have subclasses *human-agent* and *software-agent*, both of which are considered to be domain-independent.

When a new domain (such as the corporate personnel domain) is introduced to the Natural Language agent, it is usually necessary to add new classes reflecting the distinctions made in that domain. For example, the *human-agent* class might have a domain-specific subclass *employee* that is broken into subclasses *manager*, *salesperson*, *researcher*, and *programmer* — reflecting the personnel structure of a particular organization. (These are some of the classes used in our office assistant domain.)

---

[5] Interagent Communication Language — Natural Language Linker

Because the class hierarchy is so central to the expression of the rules used by the Natural Language agent, it must be easy to understand and to edit. Thus, we have provided a Class Hierarchy Editor for browsing and modification of this hierarchy. This editor also allows drag-and-drop techniques to be used in selecting classes during operation of both the ICL-NL linker and the Word Wizard, as described later.

### 4.2.3 LEAP's ICL-NL Linker

The ICL-NL Linker acquires the knowledge needed by the Natural Language agent so that it can include a new solvable (capability specification) in the formal representations that it generates from English requests.

Two main types of information are requested from the user. First, the user is asked to provide an overall characterization of the solvable as an Entity, Relationship, or Attribute. This means of characterizing solvables was selected because, as a standard part of database methodology, it is likely to be familiar to most developers, and also because the characterization can be used to guide the selection of rules that the Natural Language agent can use in generating appropriate calls to the solvable.

Second, the user is asked to annotate each solvable with information from the class hierarchy; this is done by associating a class with the functor and with each argument of each solvable. This operation is facilitated by the ability to drag and drop class names between the Class Hierarchy Editor and the ICL-NL Linker. Figure 2 shows the main window of the ICL-NL Linker being used in this way. In this example, the developer, who is characterizing the arguments of the solvables provided by an email agent, has just associated the first argument of the solvable *forward(Msg, Destination)* with the domain-specific class *message.*

In addition, the ICL-NL Linker provides several other utilities that are helpful in introducing new solvables to the Natural Language agent. For example, if a solvable represents a database relation, and thus can be queried for all the tuples in the relation, the ICL-NL Linker can be used to perform these queries and automatically create vocabulary entries corresponding to specific values of the relation's fields.

Before moving on to LEAP's most linguistically specialized component, it is worth noting that the functionality of its Class Editor and ICL-NL Linker can be viewed in a nonlinguistic context, that is, as a means of developing domain-specific ontologies, and giving characterizations of agents' capabilities in terms of these ontologies. These characterizations are general enough to be of use to more sophisticated facilitators and information brokers, which are currently under development for use with the OAA.

### 4.2.4 LEAP's Word Wizard

LEAP's Word Wizard acquires the knowledge needed by the Natural Language agent to understand sentences containing a particular word or phrase.

The Word Wizard's chief method of acquiring information from the user is exemplar-

latural
ided a
or also
f both

ent so
ations

asked
or At-
d part
ecause
nguage

e class
nent of
names
e main
eloper,
it, has
th the

intro-
resents
CL-NL
entries

noting
nonlin-
giving
teriza-
nation

ent to

mplar-

Figure 2: Using LEAP to link ontological classes to an agent specification.

based; that is, it asks the user questions about the correctness of specific phrases or sentences, and draws the appropriate conclusions based on the responses. This approach is based on previous work done at SRI on the TEAM project [6].

The Wizard operates by obtaining a categorization of a new word, and by gradually refining the categorization through a series of questions. Each refinement of category, in turn, determines the subsequent questions to be asked. Each question asked is used to (1) refine the categorization of the word (roughly, by identifying the important patterns it can be used in), (2) obtain some specific data needed about the word (such as the plural form of a noun), or (3) both of these operations. The questions are simple ones that do not require any expert knowledge about natural language processing.

For example, in constructing an agent that extracts information from a personnel database, the developer might want the agent to be able to answer questions containing the verb ‘occupy’, as in “Who occupies office number EJ219?”. After entering ‘occupy’ as a new verb, the developer would first be asked to identify one or more acceptable patterns of usage, from a list of available verb usage patterns. Assuming that he selects the pattern “A(n) ______ occupies a(n) ______”, he would then be asked to fill in the classes, from the class hierarchy, of the things that can be referred to in the blanked positions. (In this case, he might fill in the classes ‘employee’ and ‘office’.) Following this, LEAP would ask questions about the acceptability of different uses of ‘occupy’. For instance, the developer would be asked to say whether the following construction sounds OK: “An office is occupied by an employee”. From the answer, LEAP would know whether ‘occupy’ can be used in the passive form, and could use this information in generating the appropriate lexical entry for ‘occupy’, to be used by the Natural Language agent.

Once the final categorization for a new word is determined, the Wizard has all the information it needs to update the Natural Language agent’s knowledge base. The information gathered by the Wizard for a new word, along with related information entered previously using the Class Hierarchy Editor and the ICL-NL Linker, typically results in a large number of changes (perhaps 10 to 25 detailed updates) to the knowledge base. These updates are transparent to the user, who sees only the command structure provided by the user interface and the commonsense questions that have been presented.

#### 4.2.5 LEAP’s Pronunciation Wizard

Much of the knowledge needed by the Speech Recognition agent (such as a word’s part of speech) can be derived from the information acquired for the Natural Language agent. One type of linguistic knowledge that is used exclusively by the Speech Recognition agent is a word’s phonetic specification, the description of how it is pronounced. Even though the Speech Recognition agent incorporates a large corpus of phonetic information for ordinary words, the vocabulary used by an agent can include domain-specific terminology, names, abbreviations, and acronyms, and thus it is frequently the case that additional phonetic specifications are needed. As a simple example, our office assistant agent system might be expected to answer the spoken question “What is the extension of Adam Cheyer”, or to satisfy the request “Send a message to cheyer@ai.sri.com”.

Since the Speech Recognition agent needs to have a phonetic specification for each new word introduced to it, and since these specifications employ a fairly specialized notation, LEAP includes a Pronunciation Wizard to help the agent developer in entering these specifications. The Pronunciation Wizard operates in the background, checking each new word to see if its pronunciation is already known. When a word without a known pronunciation is encountered, it is placed on an action list, until the developer is ready to work on pronunciations. At that time, he can select a word from the action list, and the Pronunciation Wizard uses a sophisticated algorithm to generate a list of plausible phonetic specifications for the word. The developer is asked to select one of these, and also has the option to edit it. To assist in this task, the user can ask to see a phonetic specification for any other word known to the system. For instance, in selecting a phonetic specification for the name "Cheyer", it might be helpful to have a look at the specification for the rhyming word "buyer".

One other way in which the Pronunciation Wizard can be helpful, but which has not yet been implemented, is that a selected phonetic specification could be submitted to the OAA's text-to-speech support agent for audio playback.

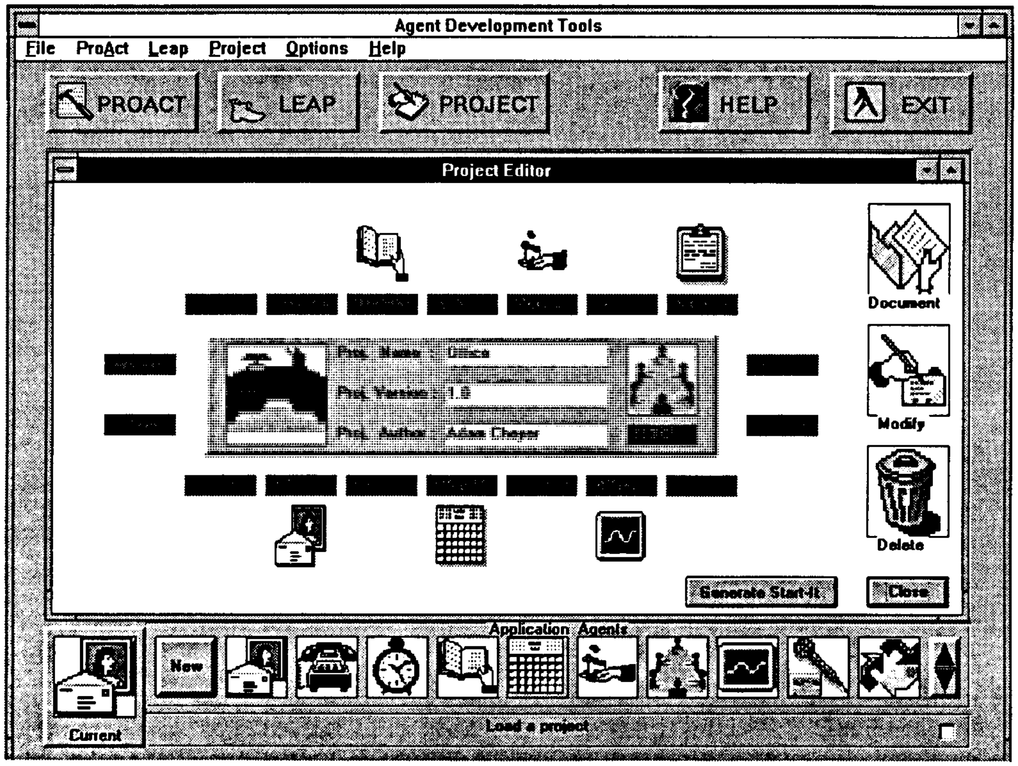
## 4.3 PROJECT: Configuring Communities of Agents

The PROJECT tool, which addresses many of the requirements expressed in Section 2.3, is used to define particular configurations of agents for a given application domain. Using PROJECT, a programmer can graphically construct an agent project by adding members to a conference table, selecting participants from repositories of available agents, and then tailoring agent execution parameters to the task at hand. These execution parameters include such things as what specific machine to execute an agent on, what facilitator the agent should connect to, and what steps to take if the agent unexpectedly crashes. Once a configuration has been specified, the PROJECT tool can generate data files for use by StartIt (Section 3.2).

In Figure 3, PROJECT's main screen is shown, with construction of a project configuration in progress.

## 5 Conclusions and Future Directions

The main theme of this paper has been that agent-based software paradigms introduce challenging new requirements for development environments, which will need to be addressed before these paradigms are able to realize their full promise. We began by identifying some important general requirements for agent development environments which are relevant to most, if not all, agent-based systems. We have outlined the architecture and functionality of one particular agent-based paradigm, the Open Agent Architecture (OAA), in order to illustrate how these general requirements arise in that context. In our presentation of the Agent Development Toolkit — a prototype development environment for OAA-based systems, which itself consists of a collection of OAA agents — we have



Proj. Name : Office

Proj. Version : 1.0

Proj. Author : Adam Cheyer

Figure 3: Using PROJECT to define an operable configuration of agents.

shown how many of these requirements have been addressed.

In building the ADT, our initial focus has been on capabilities that provide the greatest gains in productivity, and that are readily accessible to novice agent developers. We recognize that there are many possibilities for additional functionality that can be introduced into the ADT framework, and consequently we have designed the ADT for extensibility.

We have not yet taken full advantage of the fact that the ADT is itself implemented within the OAA. Thus the Natural Language and Speech Recognition agents could be used to provide a multimodal interface for the ADT, just as they have for some of our application domains. More importantly, implementation within the OAA means that the results of many development decisions can be tested immediately and demonstrated to the developer within their context of use. For example, when introducing new vocabulary for an agent using LEAP, it should be possible to immediately try out a sentence containing that vocabulary and observe, first, whether the Natural Language agent produces the correct formal representations, and second, whether these representations result in the desired set of agent interactions.

One important area that has not been addressed is debugging tools. Because of the complexity associated with interactions of multiple autonomous agents and the overhead associated with deployment on distributed sites, the ability to simulate a community of agents will have great value. We see this ability as something that will be tightly integrated with the execution environment (which again, will be facilitated by the implementation of the ADT within the OAA). For any selected configuration of agents, it should be possible to initiate a simulated set of interactions without requiring any additional setup effort. The simulation will allow for global and local views of agent activities, with the ability to inspect data, trace, set breakpoints, and step through execution.

Finally, there is important work to be done in reasoning about agent capabilities specifications. So far we have only made use of each agent's specification of the services it *provides*, but it is interesting to consider what could be done if additional information were provided by each agent as to what services it *uses*. We would like to explore to what extent, given these additional specifications, the development environment can automatically determine whether a given configuration of agents can supply a given set of services, and if not, find and select existing reusable agents that supply the missing capabilities.

## References

[1] Mic Bowman, Peter B. Danzig, Darren R. Hardy, Udi Manber, and Michael F. Schwartz. The Harvest information discovery and access system. In *Proceedings of the Second International World Wide Web Conference*, pages 763–771, Chicago, Illinois, October 1994.

[2] A. Cheyer and L. Julia. Multimodal maps: An agent-based approach. In *Proceedings of the International Conference on Cooperative Multimodal Communication*, Eindhoven, The Netherlands, May 1995.

[3] Michael Cohen, Ze'ev Rivlin, and Harry Bratt. Speech recognition in the ATIS domain using multiple knowledge sources. In *Proceedings of the ARPA Spoken Language Systems Technology Workshop*, Austin, Texas, January 1995.

[4] P. R. Cohen, A. Cheyer, M. Wang, and S. C. Baeg. An open agent architecture. In O. Etzioni, editor, *Proceedings of the AAAI Spring Symposium Series on Software Agents*, pages 1–8, Stanford, California, March 1994. American Association for Artificial Intelligence.

[5] J. Dowding, J. M. Gawron, D. Appelt, J. Bear, L. Cherny, R. Moore, and D. Moran. Gemini: A natural language system for spoken-language understanding. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 54–61, Columbus, Ohio, June 1993.

[6] Barbara J. Grosz, Douglas E. Appelt, Paul Martin, and Fernando Pereira. TEAM: An experiment in the design of transportable natural-language interfaces. Technical Note 356R, Artificial Intelligence Center, SRI International, Menlo Park, California, 1987.

[7] D. G. Schwartz. Cooperating heterogeneous systems: A blackboard-based meta approach. Technical Report 93-112, Center for Automation and Intelligent Systems Research, Case Western Reserve University, Cleveland, Ohio, April 1993. Unpublished Ph.D. thesis.

# EX. B

(Search History) ▼

OCLC NO ▼ 37324817 View Entire Collection ▼

System Sorted ▼ Search

Record: Prev Next

| | |
|---|---|
| Conference | International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (1st : 1996 : London, England) |
| Title | PAAM 96 : proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, 22nd-24th April 1996, Westminster Central Hall, London, UK |
| Published | Blackpool, Lancashire : Practical Application Co. Ltd., [1996] |

Information available at Google Books

Click here to report catalog error

| LOCATION | CALL NO. | STATUS |
|---|---|---|
| CEAS Stacks | Q335 .I585 1996 | AVAILABLE |

## Details

| | |
|---|---|
| Description | 933 p. : ill. ; 25 cm |
| Note | "Conference organization, The Practical Application Company ; programme and tutorial chairs, Barry Crabtree and Nick Jennings ; sponsorship co-ordinator, Clive Spenser." |
| | Includes bibliographical references |
| Subject | Artificial intelligence -- Congresses |
| | Distributed artificial intelligence -- Congresses |
| | Electronic data processing -- Distributed processing -- Congresses |
| Other Name | Practical Application Company |
| Other Title | Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology |
| | Practical application of intellignet agents and multi-agent technology |
| OCLC # | 37324817 |
| ISBN | 0952555433 |

Permanent link to this record: http://uclid.uc.edu/record=b2510988~S39

# Multimodal Maps: An Agent-based Approach

Adam Cheyer and Luc Julia

SRI International
333 Ravenswood Ave
Menlo Park, CA 94025 - USA

Abstract. In this paper, we discuss how multiple input modalities may be combined to produce more natural user interfaces. To illustrate this technique, we present a prototype map-based application for a travel planning domain. The application is distinguished by a synergistic combination of handwriting, gesture and speech modalities; access to existing data sources including the World Wide Web; and a mobile handheld interface. To implement the described application, a hierarchical distributed network of heterogeneous software agents was augmented by appropriate functionality for developing synergistic multimodal applications.

~~Key words: Multimodal Interface, Agent Architecture, Distributed Artificial Intelligence.~~

## 1 Introduction

As computer systems become more powerful and complex, efforts to make computer interfaces more simple and natural become increasingly important. Natural interfaces should be designed to facilitate communication in ways people are already accustomed to using. Such interfaces allow users to concentrate on the tasks they are trying to accomplish, not worry about what they must do to control the interface.

In this paper, we begin by discussing what input modalities humans are comfortable using when interacting with computers, and how these modalities should best be combined in order to produce natural interfaces. In ~~section three~~Sect. 3, we present a prototype map-based application for the travel planning domain which uses a synergistic combination of several input modalities. Section ~~four~~4 describes the agent-based approach we used to implement the application and the work on which it is based. In ~~section five~~Sect. 5, we summarize our conclusions and future directions.

## 2 Natural Input

### 2.1 Input Modalities

Direct manipulation interface technologies are currently the most widely used techniques for creating user interfaces. Through the use of menus and a graphical user interface, users are presented with sets of discrete actions and the objects on which to perform them. Pointing devices such as a mouse facilitate selection of an object or action, and drag and drop techniques allow items to be moved or combined with other entities or actions.

With the addition of electronic pen devices, gestural drawings add a new dimension direct manipulation interfaces. Gestures allow users to communicate a surprisingly wide range of meaningful requests with a few simple strokes. Research has shown that multiple gestures can be combined to form dialog, with rules of temporal grouping overriding temporal sequencing ~~[22].~~(Rhyne, 1987). Gestural commands are particularly applicable to graphical or editing type tasks.

Direct manipulation interactions possess many desirable qualities: communication is generally fast and concise; input techniques are easy to learn and remember; the user has a good idea about what can be accomplished, as the visual presentation of the available actions is generally easily accessible. However, direct manipulation suffers from limitations when trying to access or describe entities which are not or can not be visualized by the user.

Limitations of direct manipulation style interfaces can be addressed by another interface technology, that of natural language interfaces. Natural language interfaces excel in describing entities that are not currently displayed on the monitor, in specifying temporal relations between entities or actions, and in identifying members of sets. These strengths are exactly the weaknesses of direct manipulation interfaces, and concurrently, the weaknesses of natural language interfaces (ambiguity, conceptual coverage, etc.) can be overcome by the strengths of direct manipulation.

Natural language content can be entered through different input modalities, including typing, handwriting, and speech. It is important to note that, while the same textual content can be provided by the three modalities, each modality has widely varying properties.

Spoken language is the modality used first and foremost in human-human interactive problem solving ~~[4].~~(Cohen et al., 1990). Speech is an extremely fast medium, several times faster than typing or handwriting. In addition, speech input contains content that is not present in other forms of natural language input, such as prosidy, tone and characteristics of the speaker (age, sex, accent).

Typing is the most common way of entering information into a computer, because it is reasonably fast, very accurate, and requires no computational resources.

Handwriting has been shown to be useful for certain types of tasks, such as performing numerical calculations and manipulating names which are difficult to pronounce ~~[18, 19].~~(Oviatt, 1994; Oviatt and Olson, 1994). Because of its relatively slow production rate, handwriting may induce users to produce different types of input than is generated by spoken language; abbreviations, symbols and non-grammatical patterns may be expected to be more prevalent amid written input.

### 2.2 Combination of Modalities

As noted in the previous section, direct manipulation and natural language seem to be very complementary modalities. It is therefore not surprising that a number of multimodal systems combine the two.

Notable among such systems is the Cohen's Shoptalk system ~~[6],~~(Cohen, 1992), a prototype manufacturing and decision-support system that aids in tasks such as quality assurance monitoring, and production scheduling. The natural language module of Shoptalk is based on the Chat-85 natural language system ~~[25]~~(Warren and Perreira, 1982) and is particularly good at handling time, tense, and temporal reasoning.

A number of systems have focused on combining the speed of speech with the reference provided by direct manipulation of a mouse pointer. Such systems include the XTRA system ~~[1],~~(Allegayer et al, 1989), CUBRICON ~~[15],~~(Neal and Shapiro, 1991), the PAC-Amodeus model ~~[16],~~(Nigay and Coutaz, 1993), and TAPAGE ~~[9].~~(Faure and Julia, 1994).

XTRA and CUBRICON are both systems that combine complex spoken input with mouse clicks, using several knowledge sources for reference identification. CUBRICON's domain is a map-based task, making it similar to the application developed in this paper. However, the two are different in that CUBRICON can only use direct manipulation to indicate a specific item, whereas our system produces a richer mixing of modalities by adding both gestural and written language as input modalities.

The PAC-Amodeus systems such as VoicePaint and Notebook allow the user to synergistically combine vocal or mouse-click commands when interacting with notes or graphical objects. However, due to the selected domains, the natural language input is very simple, generally of the style "Insert a note here~~."~~".

TAPAGE is another system that allows true synergistic combination of spoken input with direct manipulation. Like PAC-Amodeus, TAPAGE's domain provides only simple linguistic input. However, TAPAGE uses a pen-based interface instead of a mouse, allowing gestural commands. TAPAGE, selected as a building block for our map application, will be described more in detail in ~~section~~Sect. 4.2.

Other interesting work regarding the simultaneous combination of handgestures and gaze can be found in ~~[2, 13].~~Bolt (1980) and Koons, Sparrell and Thorisson (1993).

## 3 A Multimodal Map Application

In this section, we will describe a prototype map-based application for a travel planning domain. In order to provide the most natural user interface possible, the system permits the user to simultaneously combine direct manipulation, gestural drawings, handwritten, typed and spoken natural language. When designing the system, other criteria were considered as well:

The user interface must be light and fast enough to run on a handheld PDA while able to access applications and data that may require a more powerful machine.
Existing commercial or research natural language and speech recognition systems should be used.
Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web.
As illustrated in ~~Figure~~Fig. 1, the user is presented with a pen sensitive map display on which drawn gestures and written natural language statements may be combined with spoken input. As opposed to a static paper map, the location, resolution, and content presented by the map change, according to the requests of the user. Objects of interest, such as restaurants, movie theaters, hotels, tourist sites, municipal buildings, etc. are displayed as icons. The user may ask the map to perform various actions. For example :
distance calculation : e.g. "How far is the hotel from Fisherman's Wharf?"
object location : e.g. "Where is the nearest post office?"
filtering : e.g. "Display the French restaurants within 1 mile of this hotel."
information retrieval : e.g. "Show me all available information about Alcatraz."
The application also makes use of multimodal (multimedia) output as well as input: video, text, sound and voice can all be combined when presenting an answer to a query.
During input, requests can be entered using gestures (see ~~Figure~~Fig. 2 for sample gestures), handwriting, voice, or a combination of pen and voice. For instance, in order to calculate the distance between two points on the map, a command may be issued using the following:
gesture, by simply drawing a line between the two points of interest.
voice, by speaking "What is the distance from the post office to the hotel?".
handwriting, by writing "dist p.o. to hotel?"
synergistic combination of pen and voice, by speaking "What is the distance from here to this hotel?" while simultaneously indicating the specified locations by pointing or circling.
Notice that in our example of synergistic combination of pen and voice, the arguments to the verb "distance" can be specified before, at the same time, or shortly after the vocalization of the request to calculate the distance. If a user's request is ambiguous or underspecified, the system will wait several seconds and then issue a prompt requesting additional information.
The user interface runs on pen-equipped PC's or a Dauphin handheld PDA ~~([7])~~(Dauphin, DTR-1 User's Manual) using either a microphone or a telephone for voice input. The interface is connected either by modem or ethernet to a server machine which will manage database access, natural language processing and speech recognition for the application. The result is a mobile system that provides a synergistic pen/voice interface to remote databases.
In general, the speed of the system is quite acceptable. For gestural commands, which are handled locally on the user interface machine, a response is produced in less than one second. For handwritten commands, the time to recognize the handwriting, process the English query, access a database and begin to display the results on the user interface is less than three seconds (assuming an ethernet connection, and good network and database response). Solutions to verbal commands are displayed in three to five seconds after the end of speech has been detected; partial feedback indicating the current status of the speech recognition is provided earlier.

## 4 Approach

In order to implement the application described in the previous section, we chose to augment a proven agent- based architecture with functionalities developed for a synergistically multimodal application. The result is a flexible methodology for designing and implementing distributed multimodal applications.

### 4.1 Building Blocks

~~**4.1.1**~~ Open Agent Architecture. The Open Agent Architecture (OAA) ~~[5]~~(Cohen et al., 1994) provides a framework for coordinating a society of agents which interact to solve problems for the user. Through the use of agents, the OAA provides distributed access to commercial applications, such as mail systems, calendar programs, databases, etc.
The Open Agent Architecture possesses several properties which make it a good candidate for our needs:

An Interagent Communication Language (ICL) and Query Protocol have been developed, allowing agents to communicate among themselves. Agents can run on different platforms and be implemented in a variety of programming languages.

Several natural language systems have been integrated into the OAA which convert English into the Interagent Communication Language. In addition, a speech recognition agent has been developed to provide transparent access to the Corona speech recognition system.

The agent architecture has been used to provide natural language and agent access to various heterogeneous data and knowledge sources.

Agent interaction is very fine-grained. The architecture was designed so that a number of agents can work together, when appropriate in parallel, to produce fast responses to queries.

The architecture for the OAA, based loosely on Schwartz's FLiPSiDE system~~[23],~~ (Schwartz, 1993), uses a hierarchical configuration where client agents connect to a "facilitator" server. Facilitators provide content-based message routing, global data management, and process coordination for their set of connected agents. Facilitators can, in turn, be connected as clients of other facilitators. Each facilitator records the published functionality of their sub-agents, and when queries arrive in Interagent Communication Language form, they are responsible for breaking apart any complex queries and for distributing goals to the appropriate agents. An agent solving a goal may require supporting information and the agent architecture provides numerous means of requesting data from other agents or from the user.

Among the assortment of agent architectures, the Open Agent Architecture can be most closely compared to work by the ARPA knowledge sharing community ~~[10].~~(Genesereth and Singh, 1994). The OAA's query protocol, Interagent Communication Language and Facilitator mechanisms have similar instantiations in the SHADE project, in the form of KQML, KIF and various independent capability matchmakers. Other agent architectures, such as General Magic's Telescript ~~[11],~~(General Magic, 1995), MASCOS ~~[20],~~(Park et al, submitted), or the CORBA distributed object approach ~~[17]~~(Object Management Group, 1991) do not provide as fully developed mechanisms for interagent communication and delegation.

The Open Agent Architecture provides capability for accessing distributed knowledge sources through natural language and voice, but it is lacking integration with a synergistic multimodal interface.

~~4.1.2~~ TAPAGE. TAPAGE (edition de Tableaux par la Parole et la Geste) is a synergistic pen/voice system for designing and correcting tables.

To capture signals emitted during a user's interaction, TAPAGE integrates a set of modality agents, each responsible for a very specialized kind of signal ~~[9].~~(Faure and Julia, 1994). The modality agents are connected to an ~~"interpret agent"~~'interpret agent' which is responsible for combining the inputs across all modalities to form a valid command for the application. The interpret agent receives filtered results from the modality agents, sorts the information into the correct fields, performs type-checking on the arguments, and prompts the user for any missing information, according to the model of the interaction. The interpret agent is also responsible for merging the data streams sent by the modality agents, and for resolving ambiguities among them, based on its knowledge of the application's internal state. Another function of the interpret agent is to produce reflexes: reflexes are actions output at the interface level without involving the functional core of the application.

The TAPAGE system can accept multimodal input, but it is not a distributed system; its functional core is fixed. In TAPAGE, the set of linguistic input is limited to a verb object argument format.

### 4.2 Synthesis

In the Open Agent Architecture, agents are distributed entities that can run on different machines, and communicate together to solve a task for the user. In TAPAGE, agents are used to provide streams of input to a central interpret process, responsible for merging incoming data. A generalization of these two types of agents could be:

Macro Agents: contain some knowledge and ability to reason about a domain, and can answer or make queries to other macro agents using the Interagent Communication Language.

Micro Agents: are responsible for handling a single input or output data stream, either filtering the signal to or from a hierarchically superior ~~"interpret"~~'interpret' agent.

The network architecture that we used was hierarchical at two resolutions~~–~~: micro agents are connected to a superior macro agent, and macro agents are connected in turn to a facilitator agent. In both cases, a server is responsible for the supervision of its client sub-agents.

In order to describe our implementation, we will first give a description of each agent used in our application and then illustrate the flow of communication among agents produced by a user's request.

Speech Recognition (SR) Agent: The SR agent provides a mapping from the Interagent Communication Language to the API for the Decipher (Corona) speech recognition system ~~[4],~~(Cohen et al., 1990), a continuous speech speaker independent recognizer based on Hidden Markov Model technology. This macro agent is also responsible for supervising a child micro agent whose task is to control the speech data stream. The SR agent can provide feedback to an interface agent about the current status and progress of the micro agent (e.g. "listening~~",~~", "end of speech detected", etc.) This agent is written in C.

Natural Language (NL) Parser Agent: translates English expressions into the Interagent Communication Language (ICL). For a more complete description of the ICL, see ~~[5].~~Cohen et al. (Cohen et al., 1994). The NL agent we selected for our application is the simplest of those integrated into the OAA. It is written in Prolog using Definite Clause Grammars, and supports a distributed vocabulary; each agent dynamically adds word definitions as it connects to the network. A current project is underway to integrate the Gemini natural language system ~~[4],~~(Cohen et al., 1990), a robust bottom up parser and semantic interpreter specifically designed for use in Spoken Language Understanding projects.

Database Agents: Database agents can reside at local or remote locations and can be grouped hierarchically according to content. Micro agents can be connected to database agents to monitor relevant positions or events in real time. In our travel planning application, database agents provide maps for each city, as well as icons, vocabulary and information about available hotels, restaurants, movies, theaters, municipal buildings and tourist attractions. Three types of databases were used: Prolog databases, X.500 hierarchical databases, and data loaded automatically by scanning HTML pages from the World Wide Web (WWW). In one instance, a local newspaper provides weekly updates to its Mosaic-accessible list of current movie times and reviews, as well as adding several new restaurant reviews to a growing collection; this information is extracted by an HTML reading database agent and made accessible to the agent architecture. Descriptions and addresses of new restaurants are presented to the user on request, and the user can choose to add them to the permanent database by specifying positional coordinates on the map (~~eg~~e.g. "add this new restaurant here"), information lacking in the WWW database.

Reference Resolution Agent: This agent is responsible for merging requests arriving in parallel from different modalities, and for controlling interactions between the user interface agent, database agents and modality agents. In this implementation, the reference resolution agent is domain specific: knowledge is encoded as to what actions must be performed to resolve each possible type of ICL request in its particular domain. For a given ICL logical form, the agent can verify argument types, supply default values, and resolve argument references. Some argument references are descriptive ("How far is it to the hotel on Emerson Street?"); in this case, a domain agent will try to resolve the definite reference by sending database agent requests. Other references, particularly when contextual or deictic, are resolved by the user interface agent ("What are the rates for this hotel?"). Once arguments to a query have been resolved, this~~ agent~~ agent coordinates the actions and calculations necessary to produce the result of the request.

Interface Agent: This macro agent is responsible for managing what is currently being displayed to the user, and for accepting the user's multimodal input. The Interface Agent also coordinates client modality agents and resolves ambiguities among them : handwriting and gestures are interpreted locally by micro agents and combined with results from the speech recognition agent, running on a remote speech server. The handwriting micro-agent interfaces with the Microsoft PenWindows API and accesses a handwriting recognizer by CIC Corporation. The gesture micro- agent accesses recognition algorithms developed for TAPAGE.

An important task for the interface agent is to record which objects of each type are currently salient, in order to resolve contextual references such as "the hotel" or "where I was before." Deictic references are resolved by gestural or direct manipulation commands. If no such indication is currently specified, the user interface agent waits long enough to give the user an opportunity to supply the value, and then prompts the user for it.

We shall now give an example of the distributed interaction of agents for a specific query. In the following example, all communication among agents passes transparently through a facilitator agent in an undirected fashion; this process is left out of the description for brevity.

1. A user speaks: "How far is the restaurant from this hotel?"
2. The speech recognition agent monitors the status and results from its micro agent, sending feedback received by the user interface agent. When the string is recognized, a translation is requested.
3. The English request is received by the NL agent and translated into ICL form.
4. The reference resolution agent (RR) receives the ICL distance request containing one definite and one deictic reference and asks for resolution of these references.
5. The interface agent uses contextual structures to find what "the restaurant" refers to, and waits for the user to make a gesture indicating "the hotel", issuing prompts if necessary.
6. When the references have been resolved, the domain agent (RR) sends database requests asking for the coordinates of the items in question. It then calculates the distance according to the scale of the currently displayed map, and requests the user interface to produce output displaying the result of the calculation.

## 5 Conclusions

By augmenting an existing agent-based architecture with concepts necessary for synergistic multimodal input, we were able to rapidly develop a map-based application for a travel planning task. The resulting application has met our initial requirements: a mobile, synergistic pen/voice interface providing good natural language access to heterogeneous distributed knowledge sources. The approach used was general and should provide a for developing synergistic multimodal applications for other domains.

The system described here is one of the first that accepts commands made of synergistic combinations of spoken language, handwriting and gestural input. This fusion of modalities can produce more complex interactions than in many systems and the prototype application will serve as a testbed for acquiring a better understanding of multimodal input.

In the near future, we will continue to verify and extend our approach by building other multimodal applications. We are interested in generalizing the methodology even further; work has already begun on an agent-building tool which will simplify and automate many of the details of developing new agents and domains.

# Multimodal Maps: An Agent-based Approach

*Adam CHEYER and Luc JULIA*
*SRI International*
*333 Ravenswood Ave*
*Menlo Park, CA 94025 - USA*

## Abstract:

In this paper, we discuss how multiple input modalities may be combined to produce more natural user interfaces. To illustrate this technique, we present a prototype map-based application for a travel planning domain. The application is distinguished by a synergistic combination of handwriting, gesture and speech modalities; access to existing data sources including the World Wide Web; and a mobile handheld interface. To implement the described application, a ~~hierarchical~~ distributed network of heterogeneous software agents was augmented by appropriate functionality for developing synergistic multimodal applications.

~~Key words: Multimodal Interface, Agent Architecture, Distributed Artificial Intelligence.~~

# ~~1~~ Introduction

As computer systems become more powerful and complex, efforts to make computer interfaces more simple and natural become increasingly important. Natural interfaces should be designed to facilitate communication in ways people are already accustomed to using. Such interfaces should allow users to concentrate on the tasks they are trying to accomplish, not worry about what they must do to control the interface.

In this paper, we begin by discussing what input modalities humans are comfortable using when interacting with computers, and how these modalities should best be combined in order to produce natural interfaces. In section three, we present a prototype map-based application for the travel planning domain which uses a synergistic combination of several input modalities. Section four describes the agent-based approach we used to implement the application and the work on which it is based. In section five, we summarize our conclusions and future directions.

## ~~2~~ Natural Input

### ~~2.1~~ Input Modalities

Direct manipulation interface technologies are currently the most widely used techniques for creating user interfaces. Through the use of menus and a graphical user interface, users are presented with sets of discrete actions and the objects on which to perform them. Pointing devices such as a mouse facilitate selection of an object or action, and drag and drop techniques allow items to be moved or combined with other entities or actions.

With the addition of electronic pen devices, gestural drawings add a new dimension to direct manipulation interfaces. Gestures allow users to communicate a surprisingly wide range of meaningful requests with a few simple strokes. Research has shown that multiple gestures can be combined to form dialog, with rules of temporal grouping overriding temporal sequencing ~~[22].~~[[23]]. Gestural commands are particularly applicable to graphical or editing type tasks.

Direct manipulation interactions possess many desirable qualities: communication is generally fast and concise; input techniques are easy to learn and remember; the user has a good idea about what can be accomplished, as the visual presentation of the available actions is generally easily accessible. However, direct manipulation suffers from limitations when trying to access or describe entities which are not or can not be visualized by the user.

Limitations of direct manipulation style interfaces can be addressed by another interface technology, that of natural language interfaces. Natural language interfaces excel in describing entities that are not currently displayed on the monitor, in specifying temporal relations between entities or actions, and in identifying members of sets. These strengths are exactly the weaknesses of direct manipulation interfaces, and concurrently, the weaknesses of natural language interfaces (ambiguity, conceptual coverage, etc.) can be overcome by the strengths of direct manipulation~~.~~ [[6]].

Natural language content can be entered through different input modalities, including typing, handwriting, and speech. It is important to note that, while the same textual content can be provided by the three modalities, each modality has widely varying properties.

Spoken language is the modality used first and foremost in human-human interactive problem solving ~~[~~[[4~~].~~]]. Speech is an extremely fast medium, several times faster than typing or handwriting. In addition, speech input contains content that is not present in other forms of natural language input, such as prosidy, tone and characteristics of the speaker (age, sex, accent).

Typing is the most common way of entering information into a computer, because it is reasonably fast, very accurate, and requires no computational resources.

Handwriting has been shown to be useful for certain types of tasks, such as performing numerical calculations and manipulating names which are difficult to pronounce ~~[~~[[18~~, 19].~~], [20]]. Because of its relatively slow production rate, handwriting may induce users to produce different types of input than is generated by spoken language; abbreviations, symbols and non-grammatical patterns may be expected to be more prevalent amid written input.

## ~~2.2~~ Combination of Modalities

As noted in the previous section, direct manipulation and natural language seem to be very complementary modalities. It is therefore not surprising that a number of multimodal systems combine the two.

Notable among such systems is the Cohen's Shoptalk system ~~[~~[[6~~],~~]], a prototype manufacturing and decision-support system that aids in tasks such as quality assurance monitoring, and production scheduling. The natural language module of Shoptalk is based on the Chat-85 natural language system ~~[25]~~[[26]] and is particularly good at handling time, tense, and temporal reasoning.

A number of systems have focused on combining the speed of speech with the reference provided by direct manipulation of a mouse pointer. Such systems include the XTRA system ~~[~~[[1~~],~~]], CUBRICON ~~[~~[[15~~],~~]], the PAC-Amodeus model ~~[~~[[16~~],~~]], and TAPAGE ~~[~~[[9~~].~~], [12]].

XTRA and CUBRICON are both systems that combine complex spoken input with mouse clicks, using several knowledge sources for reference identification. CUBRICON's domain is a map-based task, making it similar to the application developed in this paper. However, the two are different in that CUBRICON can only use direct manipulation to indicate a specific item, whereas our system produces a richer mixing of modalities by adding both gestural and written language as input modalities.

~~The~~ PAC-Amodeus systems such as VoicePaint and Notebook allow the user to synergistically combine vocal or mouse-click commands when interacting with notes or graphical objects. However, due in part to the selected domains, the natural language input is very simple, generally of the style "Insert a note here."

TAPAGE is another system that allows true synergistic combination of spoken input with direct manipulation. Like PAC-Amodeus, TAPAGE's domain provides only simple linguistic input. However, TAPAGE uses a pen-based interface instead of a mouse, allowing gestural commands. TAPAGE, selected as ~~a~~ one of the "building ~~block~~blocks" for our map application, will be described more in detail in section 4.2.

Other ~~interesting~~pertinent work regarding the simultaneous combination of handgestures and gaze can be found in ~~[~~[[2~~,~~ ], [13~~].~~]].

## ~~3~~ A Multimodal Map Application

In this section, we will describe a prototype map-based application for a travel planning domain. In order to provide the most natural user interface possible, the system permits the user to simultaneously combine direct manipulation, gestural drawings, handwritten, typed and spoken natural language When designing the architecture for the system, other criteria were considered as well:

The user interface must be light and fast enough to run on a handheld PDA while able to access applications and data that may require a more powerful machine.

Existing commercial or research natural language and speech recognition systems should be used.

Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML ~~form~~format on the World Wide Web.

The map functionality, interface design, and classes of input data of the system presented here is based on a design by Oviatt and Cohen, used by them in a wizard-of-oz simulation system designed to explore complex interactions of modalities [[19]]. The agent-based architecture used to realize Oviatt and Cohen's design is new, as is its application to travel planning.

As illustrated in Figure 1, the user is presented with a pen sensitive map display on which drawn gestures and ~~written~~handwritten natural language statements may be combined with spoken input. As opposed to a static paper map, the location, resolution, and content presented by the map change, according to the requests of the user. Objects of interest, such as restaurants, movie theaters, hotels, tourist sites, municipal buildings, etc. are displayed as icons. The user may ask the map to perform various actions. For example :

distance calculation : e.g. "How far is the hotel from Fisherman's Wharf?"
object location : e.g. "Where is the nearest post office?"
filtering : e.g. "Display the French restaurants within 1 mile of this hotel."
information retrieval : e.g. "Show me all available information about Alcatraz."

The application also makes use of multimodal (multimedia) output as well as input: video, text, sound and voice can all be combined when presenting an answer to a query.

During input, requests can be entered using gestures (~~see~~ Figure 2 ~~for sample gestures~~), handwriting, voice, or a combination of pen and voice. For instance, in order to calculate the distance between two points on the map, a command may be issued using the following:

gesture, by simply drawing a line between the two points of interest.
voice, by speaking "What is the distance from the post office to the hotel?".
handwriting, by writing "dist p.o. to hotel?"
synergistic combination of pen and voice, by speaking "What is the distance from here to this hotel?" while simultaneously indicating the specified locations by pointing or circling.

Notice that in our example of synergistic combination of pen and voice, the arguments to the verb "distance" can be specified before, at the same time, or shortly after the vocalization of the request to calculate the distance. If a user's request is ambiguous or underspecified, the system will wait several seconds and then issue a prompt requesting additional information.

The user interface runs on pen-equipped PC's or a Dauphin handheld PDA ~~([~~([[7~~])~~]]) using either a microphone or a telephone for voice input. The interface is connected either by modem or ethernet to a server machine which will manage database access, natural language processing and speech recognition for the application. The result is a mobile system that provides a synergistic pen/voice interface to remote databases.

In general, the speed of the system is quite acceptable. For gestural commands, which are handled locally on the user interface machine, a response is produced in less than one second. For handwritten commands, the time to recognize the handwriting, process the English query, access a database and begin to display the results on the user interface is less than three seconds (assuming an ethernet connection, and good network and database response). Solutions to verbal commands are displayed in three to five seconds after the end of speech has been detected; partial feedback indicating the current status of the speech recognition is provided earlier.

## ~~4~~ Approach

In order to implement the application described in the previous section, we chose to augment a proven agent- based architecture with functionalities developed for a synergistically multimodal application. The result is a flexible methodology for designing and implementing distributed multimodal applications.

### ~~4.1~~ Building Blocks

#### ~~4.1.1~~ Open Agent Architecture

The Open Agent Architecture (OAA) ~~[~~[[5~~]~~]] provides a framework for coordinating a society of agents which interact to solve problems for the user. Through the use of agents, the OAA provides distributed access to commercial applications, such as mail systems, calendar programs, databases, etc.

The Open Agent Architecture possesses several properties which make it a good candidate for our needs:

An Interagent Communication Language (ICL) and Query Protocol have been developed, allowing agents to communicate among themselves. Agents can run on different platforms and be implemented in a variety of programming languages.

Several natural language systems have been integrated into the OAA which convert English into the Interagent Communication Language. In addition, a speech recognition agent has been developed to provide transparent access to the Corona speech recognition system.

The agent architecture has been used to provide natural language and agent access to various heterogeneous data and knowledge sources.

Agent interaction is very fine-grained. The architecture was designed so that a number of agents can work together, when appropriate in parallel, to produce fast responses to queries.

The architecture for the OAA, based loosely on Schwartz's FLiPSiDE system~~[23],~~[[24]], uses a hierarchical configuration where client agents connect to a "facilitator" server. Facilitators provide content-based message routing, global data management, and process coordination for their set of connected agents. Facilitators can, in turn, be connected as clients of other facilitators. Each facilitator records the published functionality of their sub-agents, and when queries arrive in Interagent Communication Language form, they are responsible for breaking apart any complex queries and for distributing goals to the appropriate agents. An agent solving a goal may require supporting information and the agent architecture provides numerous means of requesting data from other agents or from the user.

Among the assortment of agent architectures, the Open Agent Architecture can be most closely compared to work by the ARPA knowledge sharing community ~~[~~[[10~~].~~]]. The OAA's query protocol, Interagent Communication Language and Facilitator mechanisms have similar instantiations in the SHADE project, in the form of KQML, KIF and various independent capability matchmakers. Other agent architectures, such as General Magic's Telescript ~~[~~[[11~~],~~]], MASCOS ~~[20],~~[[21]], or the CORBA distributed object approach ~~[~~[[17~~]~~]] do not provide as fully developed mechanisms for interagent communication and delegation.

The Open Agent Architecture provides capability for accessing distributed knowledge sources through natural language and voice, but it is lacking integration with a synergistic multimodal interface.

### ~~4.1.2~~ TAPAGE

TAPAGE (edition de Tableaux par la Parole et la Geste) is a synergistic pen/voice system for designing and correcting tables.

To capture signals emitted during a user's interaction, TAPAGE integrates a set of modality agents, each responsible for a very specialized kind of signal ~~[~~[[9~~].~~]]. The modality agents are connected to an "interpret agent" which is responsible for combining the inputs across all modalities to form a valid command for the application. The interpret agent receives filtered results from the modality agents, sorts the information into the correct fields, performs type-checking on the arguments, and prompts the user for any missing information, according to the model of the interaction. The interpret agent is also responsible for merging the data streams sent by the modality agents, and for resolving ambiguities among them, based on its knowledge of the application's internal state. Another function of the interpret agent is to produce reflexes: reflexes are actions output at the interface level without involving the functional core of the application.

The TAPAGE system can accept multimodal input, but it is not a distributed system; its functional core is fixed. In TAPAGE, the set of linguistic input is limited to a verb object argument format.

### ~~4.2~~ Synthesis

In the Open Agent Architecture, agents are distributed entities that can run on different machines, and communicate together to solve a task for the user. In TAPAGE, agents are used to provide streams of input to a central interpret process, responsible for merging incoming data. A generalization of these two types of agents could be :

Macro Agents: contain some knowledge and ability to reason about a domain, and can answer or make queries to other macro agents using the Interagent Communication Language.

Micro Agents: are responsible for handling a single input or output data stream, either filtering the signal to or from a hierarchically superior "interpret" agent.

The network architecture that we used was hierarchical at two resolutions - micro agents are connected to a superior macro agent, and macro agents are connected in turn to a facilitator agent. In both cases, a server is responsible for the supervision of its client sub-agents.

In order to describe our implementation, we will first give a description of each agent used in our application and then illustrate the flow of communication among agents produced by a user's request.

Speech Recognition (SR) Agent: The SR agent provides a mapping from the Interagent Communication Language to the API for the Decipher (Corona) speech recognition system ~~[~~[[4~~],~~]], a large vocabulary, continuous speech, speaker independent recognizer based on Hidden Markov Model technology. This macro agent is also responsible for supervising a child micro agent whose task is to control the speech data stream. The SR agent can provide feedback to an interface agent about the current status and progress of the micro agent (e.g. "listening", "end of speech detected", etc.) This agent is written in C.

Natural Language (NL) Parser Agent: translates English expressions into the Interagent Communication Language (ICL). For a more complete description of the ICL, see ~~[~~[[5~~].~~]]. The NL agent we selected for our application is the simplest of those integrated into the OAA. It is written in Prolog using Definite Clause Grammars, and supports a distributed vocabulary; each agent dynamically adds word definitions as it connects to the network. A current project is underway to integrate the Gemini natural language system ~~[4],~~[[8]], a robust bottom up parser and semantic interpreter specifically designed for use in Spoken Language Understanding projects.

Database Agents: Database agents can reside at local or remote locations and can be grouped hierarchically according to content. Micro agents can be connected to database agents to monitor relevant positions or events in real time. In our travel planning application, database agents provide maps for each city, as well as icons, vocabulary and information about available hotels, restaurants, movies, theaters, municipal buildings and tourist attractions. Three types of databases were used: Prolog databases, X.500 hierarchical databases, and data loaded automatically by scanning HTML pages from the World Wide Web (WWW). In one instance, a local newspaper provides weekly updates to its Mosaic-accessible list of current movie times and reviews, as well as adding several new restaurant reviews to a growing collection; this information is extracted by an HTML reading database agent and made accessible to the agent architecture. Descriptions and addresses of new restaurants are presented to the user on request, and the user can choose to add them to the permanent database by specifying positional coordinates on the map (eg. "add this new restaurant here"), information lacking in the WWW database.

Reference Resolution Agent: This agent is responsible for merging requests arriving in parallel from different modalities, and for controlling interactions between the user interface agent, database agents and modality agents. In this implementation, the reference resolution agent is domain specific: knowledge is encoded as to what actions must be performed to resolve each possible type of ICL request in its particular domain. For a given ICL logical form, the agent can verify argument types, supply default values, and resolve argument references. Some argument references are descriptive ("How far is it to the hotel on Emerson Street?"); in this case, a domain agent will try to resolve the definite reference by sending database agent requests. Other references, particularly when contextual or deictic, are resolved by the user interface agent

("What are the rates for this hotel?"). Once arguments to a query have been resolved, this agent agent coordinates the actions and calculations necessary to produce the result of the request.

Interface Agent: This macro agent is responsible for managing what is currently being displayed to the user, and for accepting the user's multimodal input. The Interface Agent also coordinates client modality agents and resolves ambiguities among them : handwriting and gestures are interpreted locally by micro agents and combined with results from the speech recognition agent, running on a remote speech server. The handwriting micro-agent interfaces with the Microsoft PenWindows API and accesses a handwriting recognizer by CIC Corporation. The gesture micro- agent accesses recognition algorithms developed for TAPAGE.

An important task for the interface agent is to record which objects of each type are currently salient, in order to resolve contextual references such as "the hotel" or "where I was before." Deictic references are resolved by gestural or direct manipulation commands. If no such indication is currently specified, the user interface agent waits long enough to give the user an opportunity to supply the value, and then prompts the user for it.

We shall now give an example of the distributed interaction of agents for a specific query. In the following example, all communication among agents passes transparently through a facilitator agent in an undirected fashion; this process is left out of the description for brevity.

1. A user speaks: "How far is the restaurant from this hotel?"
2. The speech recognition agent monitors the status and results from its micro agent, sending feedback received by the user interface agent. When the string is recognized, a translation is requested.
3. The English request is received by the NL agent and translated into ICL form.
4. The reference resolution agent (RR) receives the ICL distance request containing one definite and one deictic reference and asks for resolution of these references.
5. The interface agent uses contextual structures to find what "the restaurant" refers to, and waits for the user to make a gesture indicating "the hotel", issuing prompts if necessary.
6. When the references have been resolved, the domain agent (RR) sends database requests asking for the coordinates of the items in question. It then calculates the distance according to the scale of the currently displayed map, and requests the user interface to produce output displaying the result of the calculation.

## ~~5~~ CONCLUSIONS

By augmenting an existing agent-based architecture with concepts necessary for synergistic multimodal input, we were able to rapidly develop a map-based application for a travel planning task. The resulting application has met our initial requirements: a mobile, synergistic pen/voice interface providing good natural language access to heterogeneous distributed knowledge sources. The approach used was general and should provide a means for developing synergistic multimodal applications for other domains.

The system described here is one of the first that accepts commands made of synergistic combinations of spoken language, handwriting and gestural input. This fusion of modalities can produce more complex interactions than in many systems and the prototype application will serve as a testbed for acquiring a ~~better~~deeper understanding of multimodal input.

In the near future, we will continue to verify and extend our approach by building other multimodal applications. We are interested in generalizing the methodology ~~even~~ further; work has already begun on an agent-building tool which will simplify and automate many of the details of developing new agents and domains.

## Acknowledgements

The work reported here would not have been possible without the inspiration of Sharon Oviatt and Phil Cohen under whose direction we worked for a year on a project (NSF Grant No. IRI-9213472) in which the combination of modalities contained in the interface presented here was crystallized and studied via simulations. Neither they nor their sponsors, of course, are responsible for the work presented here.

PARK
IP TRANSLATIONS
welocalize

December 18, 2017

Certification

Park IP Translations

TRANSLATOR'S DECLARATION:

I, Linda Henson, hereby declare:

That I possess advanced knowledge of the Dutch and English languages. The attached Dutch into English translation has been translated by me and to the best of my knowledge and belief, it is a true and accurate translation of the documents titled:

- Ex C internal records
- Ex D internal library catalog entry

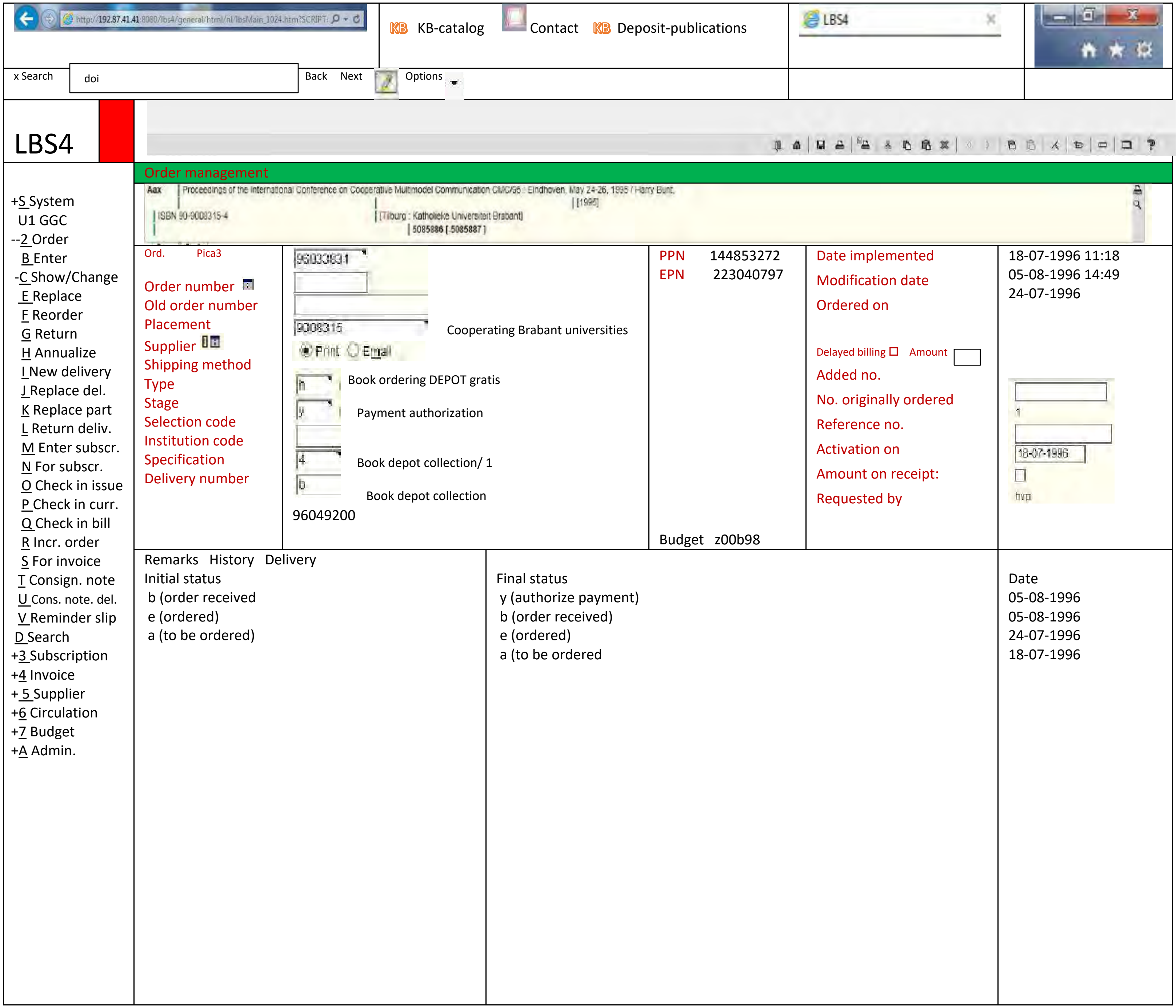
Linda R. Henson

Linda Henson

Project Number: BBLLP_1712_050

15 W. 37th Street 8th Floor
New York, NY 10018
212.581.8870
ParkIP.com

EX. C



doi

LBS4

Order management

Aax | Proceedings of the International Conference on Cooperative Multimodel Communication CMC/95 : Eindhoven, May 24-26, 1995 / Harry Bunt.
| [1995]
| ISBN 90-9008315-4
| [Tilburg : Katholieke Universiteit Brabant]
| 5085886 [ 5085887 ]

Ord. Pica3

Order number: 96033831
Old order number:
Placement:
Supplier: 9008315 Cooperating Brabant universities
● Print ○ Email
Shipping method: h Book ordering DEPOT gratis
Type: y Payment authorization
Stage:
Selection code: 4 Book depot collection/ 1
Institution code: b Book depot collection
Specification:
Delivery number: 96049200

PPN 144853272
EPN 223040797

Budget z00b98

Date implemented: 18-07-1996 11:18
Modification date: 05-08-1996 14:49
Ordered on: 24-07-1996

Delayed billing ☐ Amount

Added no.:
No. originally ordered: 1
Reference no.:
Activation on: 18-07-1996
Amount on receipt: ☐
Requested by: hvp

Remarks History Delivery

| Initial status | Final status | Date |
|---|---|---|
| b (order received | y (authorize payment) | 05-08-1996 |
| e (ordered) | b (order received) | 05-08-1996 |
| a (to be ordered) | e (ordered) | 24-07-1996 |
|  | a (to be ordered | 18-07-1996 |

EX. D

```
SET: S3 [54] TTL: 32 PPN: 144853272 PAG: 1.
Implemented: 1001:23-01-96 Amended: 1999:22-08-13 06:39:29 Status: 9999:99-
99-99
0500 Aax
0501 #tekst=txt %%rdacontent/dut
0502 #zonder medium=n %%rdamedia/dut
0503 #band=nc %%rdacarrier/dut
1100 1995 $ [1995]
1121 u
1500 /1eng
1700 /1nl
2000 9090083154
2020 B9635959
2097 #OCoLC#69071749
3011 Harry@Bunt!068920075!Hendrik Cornelis Bunt (1944-) (ISNI 0000 0001 2149
0086)
3012 Robbert-Jan@Beun!075105888!Robbert-Jan Beun (ISNI 0000 0000 8317 9093)
3161 @International Conference on Cooperative Multimodel Communication CMC/95
(Eindhoven): 1995
4000 @Proceedings of the International Conference on Cooperative Multimodel
Communication CMC/95: Eindhoven, May 24-26, 1995 / Harry Bunt, Robbert-Jan
Beun &
Tijn Borghuis (eds.)
4030 [Tilburg: Katholieke Universiteit Brabant]
4031 [Eindhoven: Technische Universiteit Eindhoven]
4060 2 dl. (VII, 324 p)
4061 ill
4062 30 cm
4204 Met lit. opg., reg
5201 !12160800X!multimedia
5202 !075635143!communicatie
5203 !075603195!computertoepassingen
3521 !075385899!@Katholieke Universiteit Brabant, Tilburg
3522 !075382903!@Technische Universiteit Eindhoven
4701 ea
4900 13-09-96 13:53:50.671
7001 13-09-96: gdfg
7100 5085886 [-5085887] !d! @ f
8008 rp/29
8009 rp/32
7900 18-09-96 14:29:57.169
7800 223040797
```

Paper No. 4
Filed: January 26, 2018

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

DISH NETWORK CORPORATION AND DISH NETWORK L.L.C.

Petitioners

v.

IPA TECHNOLOGIES INC.

Patent Owner

---

Case No. IPR2018-00351
U.S. Patent No. 6,757,718
FILED: JUNE 30, 2000
ISSUED: JUNE 10, 2004
INVENTORS: CHRISTINE HALVERSON ET AL.

TITLE: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

---

# PATENT OWNER'S MANDATORY NOTICES
## 37 C.F.R. 42.8(a)(2)

Pursuant to 37 C.F.R. § 42.8(a)(2), Patent Owner submits the following mandatory notices:

**(1) Real Party-in-interest**

The real party-in-interest is the Patent Owner, IPA Technologies Inc., which is a wholly owned subsidiary of Wi-LAN Technologies Inc. (a Delaware corporation), which is a wholly owned subsidiary of Wi-LAN Inc. (a Canadian corporation), which is a wholly owned subsidiary of Quarterhill Inc. (a Canadian corporation publicly traded on the TSX and NASDAQ).

**(2) Related matters**

Pursuant to 37 C.F.R. § 42.8(b)(2), Patent Owner submits that the '718 patent is involved in the following proceedings:

*DISH Network Corporation, et al. v. IPA Technologies Inc.*, IPR2018-00351 (PTAB); *Google LLC v. IPA Technologies Inc.*, IPR2018-00476 (PTAB); *IPA Technologies, Inc. v. DISH Network Corporation, et al.* No. 1:16-CV-01170 (D. Del.); *IPA Technologies Inc. v. NVIDIA Corporation.,* No. 1-17-cv-00287 (D. Del.); *IPA Technologies Inc. v. Sony Electronics Inc., et al.,* No. 1-17-cv-00055 (D. Del.); *IPA Technologies Inc. v. Amazon.com, Inc. et al.,* No. 1-16-cv-01266 (D. Del.).

**(3) Lead and back-up counsel**

Patent Owner provides the following designation and service information for lead and back-up counsel. 37 C.F.R. § 42.8(b)(3) and (b)(4). Please direct all correspondence regarding this proceeding to lead and back-up counsel at their respective email addresses listed below. 37 C.F.R. § 42.8(b)(4).

| LEAD COUNSEL | BACK-UP COUNSEL |
|---|---|
| **Steven W. Hartsell**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com<br>Registration No: 58,788 | **Alexander E. Gasser**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com<br>Registration No: 48,760 |
|  | **Sarah E. Spires**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com<br>Registration No: 61,501 |

January 26, 2018

Respectfully Submitted,

*/Steven W. Hartsell/*
Lead Counsel for Patent Owner
Reg. No. 58,788
**SKIERMONT DERBY LLP**
1601 Elm Street, Suite 4400
Dallas, Texas 75201
P: 214-978-6600/F: 214-978-6601

# CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the foregoing Patent Owner's Mandatory Notices were served on January 26, 2018, by delivering a copy via electronic mail to the attorneys of record for the Petitioners as follows:

Eliot Williams
G. Hopkins Guy
Ali Dhanani
BAKER BOTTS L.L.P.
eliot.williams@bakerbotts.com
hop.guy@bakerbotts.com
ali.dhanani@bakerbotts.com

Dated: January 26, 2018

Respectfully submitted,

*/Steven W. Hartsell/*
Lead Counsel for Patent Owner
Reg. No. 58,788
**SKIERMONT DERBY LLP**
1601 Elm Street, Suite 4400
Dallas, Texas 75201
P: 214-978-6600/F: 214-978-6601

Paper No. 3
Filed: January 26, 2018

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

DISH NETWORK CORPORATION AND DISH NETWORK L.L.C.

Petitioners

v.

IPA TECHNOLOGIES INC.

Patent Owner

Case No. IPR2018-00351
U.S. Patent No. 6,757,718
FILED: JUNE 30, 2000
ISSUED: JUNE 10, 2004
INVENTORS: CHRISTINE HALVERSON ET AL.

TITLE: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

# POWER OF ATTORNEY

Pursuant to 37 C.F.R. § 42.10(b), the Patent Owner of U.S. Patent No. 6,757,718, IPA Technologies Inc., hereby appoints the counsel identified below as its attorneys to transact all business in the United States Patent & Trademark Office associated with this *Inter Partes* review of U.S. Patent No. 6,757,718:

| LEAD COUNSEL | BACK-UP COUNSEL |
|---|---|
| **Steven W. Hartsell**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com<br>Reg. No. 58,788 | **Sarah E. Spires**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com)<br>Reg. No. 61,501 |
|  | **Alexander E. Gasser**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com<br>Reg. No. 48,760 |
|  | **Paul J. Skiermont**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com<br>(*pro hac vice* application to be submitted) |
|  | **Sadaf R. Abdullah**<br>SKIERMONT DERBY LLP<br>1601 Elm Street, Suite 4400<br>Dallas, Texas 75201<br>Tel: (214) 978-6600<br>Fax: (214) 978-6601<br>IPA_SDTeam@skiermontderby.com |

| | |
|---|---|
| | (*pro hac vice* application to be submitted) |
| | **Mieke K. Malmberg**<br>SKIERMONT DERBY LLP<br>800 Wilshire Boulevard, Suite 1450<br>Los Angeles, CA 90017<br>Tel: (213) 788-4500<br>Fax: (213) 788-4545<br>IPA_SDTeam@skiermontderby.com<br>(*pro hac vice* application to be submitted) |

The individual signing below has the authority to execute this document on behalf of Patent Owner, IPA Technologies Inc.

SIGNATUR :

NAME: Michael Zhang

TITLE: Director, Business Development

DATE: January 26, 2018

# CERTIFICATE OF SERVICE

The undersigned hereby certifies that a copy of the foregoing Patent Owner's Power of Attorney was served on January 26, 2018, by delivering a copy via electronic mail to the attorneys of record for the Petitioners as follows:

Eliot Williams
G. Hopkins Guy
Ali Dhanani
BAKER BOTTS L.L.P.
eliot.williams@bakerbotts.com
hop.guy@bakerbotts.com
ali.dhanani@bakerbotts.com

Dated: January 26, 2018

Respectfully submitted,

*/Steven W. Hartsell/*
Counsel for Patent Owner
**SKIERMONT DERBY LLP**
1601 Elm Street, Suite 4400
Dallas, Texas 75201
P: 214-978-6600/F: 214-978-6601

Trials@uspto.gov
571-272-7822

Paper No. 5

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

DISH NETWORK CORPORATION AND DISH NETWORK, L.L.C.,
Petitioner,

v.

IPA TECHNOLOGIES, INC.,
Patent Owner.

Case IPR2018-00351
Patent 6,757,718

Mailed: February 7, 2018

Before Amy Kattula, *Trial Paralegal*

# NOTICE OF FILING DATE ACCORDED TO PETITION AND TIME FOR FILING PATENT OWNER PRELIMINARY RESPONSE

The petition for *inter partes* review filed in the above proceeding has been accorded the filing date of December 20, 2017.

Patent Owner may file a preliminary response to the petition no later than three months from the date of this notice. The preliminary response is limited to setting forth the reasons why the requested review should not be instituted. Patent Owner may also file an election to waive the preliminary response to expedite the proceeding. For more information, please consult the Office Patent Trial Practice Guide, 77 Fed. Reg. 48756 (Aug. 14, 2012), which is available on the Board Web site at http://www.uspto.gov/PTAB.

Patent Owner is advised of the requirement to submit mandatory notice information under 37 C.F.R. § 42.8(a)(2) within 21 days of service of the petition.

The parties are encouraged to use the heading on the first page of this Notice for all future filings in the proceeding.

The parties are advised that under 37 C.F.R. § 42.10(c), recognition of counsel *pro hac vice* requires a showing of good cause. The parties are authorized to file motions for *pro hac vice* admission under 37 C.F.R. § 42.10(c). Such motions shall be filed in accordance with the "Order -- Authorizing Motion *for Pro Hac Vice* Admission" in Case IPR2013-00639, Paper 7, a copy of which is available on the Board Web site under "Representative Orders, Decisions, and Notices."

The parties are reminded that unless otherwise permitted by 37 C.F.R. § 42.6(b)(2), all filings in this proceeding must be made electronically in Patent Trial and Appeal Board End to End (PTAB E2E), accessible from the Board Web site at http://www.uspto.gov/PTAB. To file documents, users must register with PTAB E2E. Information regarding how to register with and use PTAB E2E is available at the Board Web site.

If there are any questions pertaining to this notice, please contact Amy Kattula at 571-272-5826 or the Patent Trial and Appeal Board at 571-272-7822.

PETITIONER:

Eliot Williams
Eliot.williams@bakerbotts.com

Hopkins Guy
Hop.guy@bakerbotts.com

Ali Dhanani
Ali.dhanani@bakerbotts.com

PATENT OWNER:

Steven Hartsell
shartsell@skiermontderby.com

Alexander Gasser
agasser@skiermontderby.com

Sarah Spires
sspires@skiermontderby.com

## NOTICE CONCERNING ALTERNATIVE DISPUTE RESOLUTION (ADR)

The Patent Trial and Appeal Board (PTAB) strongly encourages parties who are considering settlement to consider alternative dispute resolution as a means of settling the issues that may be raised in an AIA trial proceeding. Many AIA trials are settled prior to a Final Written Decision. Those considering settlement may wish to consider alternative dispute resolution techniques early in a proceeding to produce a quicker, mutually agreeable resolution of a dispute or to at least narrow the scope of matters in dispute. Alternative dispute resolution has the potential to save parties time and money.

Many non-profit organizations, both inside and outside the intellectual property field, offer alternative dispute resolution services. Listed below are the names and addresses of several such organizations. The listings are provided for the convenience of parties involved in cases before the PTAB; the PTAB does not sponsor or endorse any particular organization's alternative dispute resolution services. In addition, consideration may be given to utilizing independent alternative dispute resolution firms. Such firms may be located through a standard keyword Internet search.

| CPR INSTITUTE FOR DISPUTE RESOLUTION | AMERICAN INTELLECTUAL PROPERTY LAW ASSOCIATION (AIPLA) | AMERICAN ARBITRATION ASSOCIATION (AAA) | WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO) | AMERICAN BAR ASSOCIATION (ABA) |
|---|---|---|---|---|
| Telephone: (212) 949-6490 | Telephone: (703) 415-0780 | Telephone: (212) 484-3266 | Telephone: 41 22 338 9111 | Telephone : (202) 662-1000 |
| Fax: (212) 949-8859 | Fax: (703) 415-0786 | Fax: (212) 307-4387 | Fax: 41 22 733 5428 | N/A |
| 575 Lexington Ave | 241 18th Street, South, Suite 700 | 140 West 51st Street | 34, chemin des Colombettes | 1050 Connecticut Ave, NW |
| New York, NY 10022 | Arlington, VA 22202 | New York, NY 10020 | CH-1211 Geneva 20, Switzerland | Washington D.C. 20036 |
| www.cpradr.org | www.aipla.org | www.adr.org | www.wipo.int | www.americanbar.org |

If parties to an AIA trial proceeding consider using alternative dispute resolution, the PTAB would like to know whether the parties ultimately decided to engage in alternative dispute resolution and the reasons why or why not. If the parties actually engage in alternative dispute resolution, the PTAB would be interested to learn what mechanism (e.g., arbitration, mediation, etc.) was used and the general result. Such a statement from the parties is not required but would be helpful to the PTAB in assessing the value of alternative dispute resolution to parties involved in AIA trial proceedings. To report an experience with ADR, please forward a summary of the particulars to the following email address: PTAB_ADR_Comments@uspto.gov

# File History Content Report

The following content is missing from the original file history record obtained from the United States Patent and Trademark Office. No additional information is available.

Document Date - 2017-12-21

Document Title - Petition Re:

Additional Comments Requesting Trial

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

GOOGLE LLC
Petitioner

v.

IPA TECHNOLOGIES INC.
Patent Owner

---

Patent No. 6,757,718

---

# POWER OF ATTORNEY FOR PETITIONER

Pursuant to 37 C.F.R. § 42.10(b), Google LLC hereby revokes any previous powers of attorney given in this proceeding and hereby appoints the practitioners associated with Paul Hastings LLP, Customer Number 36,183, including Naveen Modi, Daniel Zeilberger, and Arvind Jairam as its attorneys to transact all business before the Patent Trial and Appeal Board of the United States Patent & Trademark Office in connection with all *inter partes* review proceedings involving U.S. Patent No. 6,757,718. Counsel's contact and service information is provided below:

| Lead Counsel |
|---|
| Naveen Modi (Reg. No. 46,224)<br>Paul Hastings LLP<br>875 15th Street NW<br>Washington, DC 20005<br>Telephone: (202) 551-1990<br>Facsimile: (202) 551-0490<br>E-mail: PH-Google-IPA-IPR@paulhastings.com |
| Back-Up Counsel |
| Daniel Zeilberger (Reg. No. 65,349)<br>Paul Hastings LLP<br>875 15th Street NW<br>Washington, DC 20005<br>Telephone: (202) 551-1993<br>Facsimile: (202) 551-0493<br>E-mail: PH-Google-IPA-IPR@paulhastings.com |
| Back-Up Counsel |
| Arvind Jairam (Reg. No. 62,759)<br>Paul Hastings LLP<br>875 15th Street NW<br>Washington, DC 20005<br>Telephone: (202) 551-1887<br>Facsimile: (202) 551-0387<br>E-mail: PH-Google-IPA-IPR@paulhastings.com |

Dated: December 21, 2017 By: __________________________________

Renny Hwang
Director, Litigation

# CERTIFICATE OF SERVICE

I hereby certify that on January 12, 2018, I caused a true and correct copy of the foregoing Power of Attorney for Petitioner to be served via express mail on the Patent Owner at the following correspondence address of record as listed on PAIR:

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

By: /Naveen Modi/
Naveen Modi (Reg. No. 46,224)

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

GOOGLE LLC
Petitioner

v.

IPA TECHNOLOGIES INC.
Patent Owner

---

Patent No. 6,757,718

---

# PETITION FOR *INTER PARTES* REVIEW OF U.S. PATENT NO. 6,757,718

## TABLE OF CONTENTS

## TABLE OF AUTHORITIES

**Page(s)**

### Federal Cases

### Federal Statutes

**Regulations**

## LIST OF EXHIBITS

| | |
|---|---|
| Ex. 1001 | U.S. Patent No. 6,757,718 |
| Ex. 1002 | Declaration of Dr. Dan R. Olsen Jr. |
| Ex. 1003 | Curriculum Vitae of Dr. Dan R. Olsen Jr. |
| Ex. 1004 | Prosecution History of U.S. Patent No. 6,757,718 |
| Ex. 1005 | U.S. Patent No. 6,742,021 |
| Ex. 1006 | Prosecution History of U.S. Patent No. 6,742,021 |
| Ex. 1007 | U.S. Patent No. 6,851,115 |
| Ex. 1008 | Prosecution History of U.S. Patent No. 6,851,115 |
| Ex. 1009 | U.S. Provisional Application No. 60/124,718 |
| Ex. 1010 | U.S. Provisional Application No. 60/124,719 |
| Ex. 1011 | U.S. Provisional Application No. 60/124,720 |
| Ex. 1012 | Cheyer *et al.*, "Multimodal Maps: An Agent-based Approach" ("*Cheyer*") |
| Ex. 1013 | U.S. Patent No. 5,197,005 to Shwartz *et al.* ("*Shwartz*") |
| Ex. 1014 | U.S. Patent No. 5,748,974 to Johnson ("*Johnson*") |
| Ex. 1015 | U.S. Patent No. 6,188,985 to Thrift *et al.* ("*Thrift*") |
| Ex. 1016 | U.S. Patent No. 6,345,389 to Dureau ("*Dureau*") |
| Ex. 1017 | U.S. Patent No. 5,841,431 to Simmers ("*Simmers*") |
| Ex. 1018 | U.S. Patent No. 6,035,197 to Haberman *et al.* ("*Haberman*") |
| Ex. 1019 | Letter from IPA Technologies, Inc.'s litigation counsel to Judge Andrews regarding claim construction in related district court litigation |

| | |
|---|---|
| Ex. 1020 | Coen, M. H., "Building Brains for Rooms: Designing Distributed Software Agents," AAAI'97/IAAI'97 *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence* (1997) ("*Coen*") |
| Ex. 1021 | Hodjat *et al.*, "An adaptive agent oriented software architecture," in Lee *et al.* (eds.) PRICAI'98: Topics in Artificial Intelligence, *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, vol. 1531, Springer, Berlin, Heidelberg (1998) ("*Hodjat*") |
| Ex. 1022 | U.S. Patent No. 5,584,024 to Shwartz ("*Shwartz*") |
| Ex. 1023 | Cheyer *et al.*, "MVIEWS: Multimodal Tools for the Video Analyst," in *Proceedings of the 1998 International Conference on Intelligent User Interfaces (IUI '98)*, San Francisco, California (Jan. 1998) |
| Ex. 1024 | Kehler *et al.*, "On Representing Salience and Reference in Multimodal Human-Computer Interaction," in *Proceedings of AAAI 1998 workshop on Representations for Multi-Modal Human-Computer Interaction*, Madison, Wisconsin (1998) |
| Ex. 1025 | Cohen *et al.*, "An Open Agent Architecture," in *Proceedings AAAI Spring Symposium*, Stanford, California (March 1994) ("*Cohen*") |
| Ex. 1026 | Martin *et al.*, "Information brokering in an agent architecture," in *Proceedings of the Second International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, Blackpool, Lancashire, UK (Apr. 1997) ("*Martin*") |
| Ex. 1027 | Wyard *et al.*, "Spoken language systems – beyond prompt and response," *BT Technol. J.* Vol. 14 No. 1 (Jan. 1996) ("*Wyard*") |
| Ex. 1028 | Excerpts from Knaster, B., Presenting Magic Cap, A Guide to General Magic's Revolutionary Communicator Software, 1994 |

| | |
|---|---|
| Ex. 1029 | Moran *et al.*, "Multimodal User Interfaces in the Open Agent Architecture," *Proc. of the 2nd International Conference on Intelligent User Interfaces (IUI '97)*, Orlando, Florida (1997) ("*Moran*") |
| Ex. 1030 | Archived copy from 1997 of SRI website http:/www.ai.sri.com/~Cheyer/mmap.html from https://web.archive.org |
| Ex. 1031 | Archived copy from 1997 of SRI website http://www.ai.sri.com:80/~Cheyer/papers/mmap/mmap.html from https://web.archive.org) |
| Ex. 1032 | Excerpts from Bunt, H., *et al.* (eds.), Multimodal Human-Computer Communication: Systems, Techniques, and Experiments, *Lecture Notes in Artificial Intelligence* 134 (Springer, copyright 1998) |
| Ex. 1033 | Konstan, J. A., "State Problems in Programming Human-Controlled Devices," *IEEE Transactions on Consumer Electronics*, vol. 40, no. 4 (Nov. 1994) ("*Konstan*") |

# I. INTRODUCTION

Google LLC ("Petitioner") requests *inter partes* review ("IPR") of claims 1-27 ("the challenged claims") of U.S. Patent No. 6,757,718 ("the '718 patent") (Ex. 1001), which, according to PTO records, is assigned to IPA Technologies Inc. ("Patent Owner"). For the reasons discussed below, the challenged claims should be found unpatentable and canceled.

# II. MANDATORY NOTICES

**Real Parties-in-Interest**: Petitioner identifies Google LLC as the real party-in-interest.

**Related Matters**: The '718 patent is at issue in the following cases: *IPA Technologies Inc. v. NVIDIA Corporation*, Case No. 1-17-cv-00287 (D. Del.), *IPA Technologies Inc. v. Sony Electronics Inc.*, Case No. 1-17-cv-00055 (D. Del.), *IPA Technologies Inc. v. Amazon.com, Inc.*, Case No. 1-16-cv-01266 (D. Del.), *IPA Technologies Inc. v. DISH Network Corporation*, Case No. 1-16-cv-01170 (D. Del.), *DISH Network Corporation et al v. IPA Technologies Inc.*, IPR2018-00351 (PTAB).

**Counsel and Service Information**: Lead counsel: Naveen Modi (Reg. No. 46,224). Backup counsel: (1) Daniel Zeilberger (Reg. No. 65,349), and (2) Arvind Jairam (Reg. No. 62,759). Service information is Paul Hastings LLP, 875 15th St.

N.W., Washington, D.C., 20005, Tel.: 202.551.1700, Fax: 202.551.1705, email: PH-Google-IPA-IPR@paulhastings.com. Petitioner consents to electronic service.

## III. PAYMENT OF FEES

The PTO is authorized to charge any fees due during this proceeding to Deposit Account No. 50-2613.

## IV. GROUNDS FOR STANDING

Petitioner certifies that the ’718 patent is available for IPR and Petitioner is not barred or estopped from requesting IPR on the grounds identified herein.

## V. PRECISE RELIEF REQUESTED AND GROUNDS RAISED

The challenged claims should be canceled as unpatentable based on the following grounds:

**Ground 1**: Claims 1-4, 6, 8-10, 12, 13, 15, 17-19, 21, 22, 24, 26, and 27 are unpatentable under pre-AIA 35 U.S.C. § 103 based on *Cheyer* (Ex. 1012) in view of *Shwartz* (Ex. 1013) and *Thrift* (Ex. 1015);

**Ground 2**: Claims 2, 11, and 20 are unpatentable under § 103 based on *Cheyer* in view of *Shwartz*, *Thrift*, and *Dureau* (Ex. 1016);

**Ground 3**: Claims 4, 13, and 22 are unpatentable under § 103 based on *Cheyer* in view of *Shwartz*, *Thrift*, and *Johnson* (Ex. 1014); and

**Ground 4**: Claims 5, 7, 14, 16, 23, and 25 are unpatentable under § 103 based on *Cheyer* in view of *Shwartz*, *Thrift*, and *Simmers* (Ex. 1017).

For purposes of this proceeding only, Petitioner assumes the earliest effective filing date of the ’718 patent is March 17, 1999, which is the filing date of three provisional applications to which the ’718 patent claims priority. (Ex. 1001, Cover.)

*Cheyer* was published several times years before the earliest effective filing date of the ’718 patent, and thus qualifies as prior art under pre-AIA 35 U.S.C. § 102(b). *Cheyer* itself has a June 1995 date on its first page. (Ex. 1012, 1.) However, *Cheyer* was actually initially published in May 1995 at the First International Conference on Cooperative Multimodal Communication (CMC/95). For example, a later book intended to document the papers released at the May 1995 conference, which book itself was published by no later than May 15, 1998 (Ex. 1032, 5 (stamp)), and itself includes a version of *Cheyer* (with minor revisions) (*id.*, 9-19), indicates that *Cheyer* was published in 1995 at the CMC/95 conference. (*Id.*, 6 (Preface).)

In any event, there is little question that *Cheyer* was widely available more than a year before the earliest effective filing date of the ’718 patent. For example, a paper by Moran *et al.* (Ex. 1029) published in 1997 (*id.*, 1, 2), includes a citation to *Cheyer*, (*id.*, 10), and in fact includes instructions on how to retrieve *Cheyer* (*id.*, 68 (“Also http://www.ai.sri.com/~oaa/ + ‘Bibliography’”)).

Similarly, a web page of the original assignee SRI International ("SRI") (http:/www.ai.sri.com/~Cheyer/mmap.html), archived by the Internet Archive, describes *Cheyer* with respect to the CMC/95 conference, specifies "24-26 May 1995" as the date, and includes a link to download *Cheyer*. (Ex. 1030, 1.) The URL of the Internet Archive page (*id.*) shows that the web page was available in 1997. *See SDI Techs., Inc. v. Bose Corp.*, IPR2014-00343, Paper No. 32 at 14 (June 11, 2015); *see also id.*, 12-17. Indeed, a full viewable copy of *Cheyer* was made available at SRI's website at least as early as 1997. (Ex. 1031, 1-22 (URL at bottom of each page shows the web pages were archived in 1997).) Thus, *Cheyer* was publicly disseminated at the CMC/95 conference in 1995 and was in any event made available on the SRI website by at least 1997.

*Shwartz* issued on March 23, 1993. Therefore, *Shwartz* is prior art at least under § 102(b).

*Thrift* was filed on October 3, 1997 and issued on February 13, 2001. Therefore, *Thrift* is available as prior art at least under § 102(e).

*Dureau* issued February 5, 2002 from U.S. Patent Application No. 09/176,611 filed October 21, 1998. Therefore, *Dureau* is available as prior art at least under § 102(e)

*Johnson* was filed on December 13, 1994 and issued on May 5, 1998. Therefore, *Johnson* is prior art at least under §§ 102 (a) and (e).

*Simmers* issued November 24, 1998 and was filed November 15, 1996 and is thus available as prior art at least under §§ 102(a) and (e).

*Thrift*, *Dureau*, and *Simmers* were not considered by the Patent Office during prosecution of the ’718 patent. (*See*, *e.g.*, Ex. 1001; Ex. 1004.) *Cheyer* was cited in an Information Disclosure Statement for a related application (Ex. 1008, 330), and *Johnson* and *Shwartz* were cited in an Information Disclosure Statement during prosecution of the ’718 patent (Ex. 1004, 83-84). However, the Examiner did not cite any of these references in any claim rejections, and Petitioner presents them in a new light never considered by the Patent Office and supported by new expert testimony (Ex. 1002). In particular, *Cheyer*, *Johnson*, and *Shwartz* are presented as part of obviousness combinations that have not been previously considered by the Patent Office.

## VI. LEVEL OF ORDINARY SKILL

A person of ordinary skill in the art as of the claimed priority date of the ’718 patent (“POSITA”) would have had at least a Bachelor’s degree in computer science, electrical engineering, or a similar discipline, and one to two years of work experience in user interfaces for computer systems (including speech-based interfaces), networked computer systems, or a related area. (Ex. 1002, ¶¶14-15.)[1] More education can substitute for practical experience and *vice versa*. (*Id.*)

## VII. OVERVIEW OF THE ’718 PATENT AND THE PRIOR ART

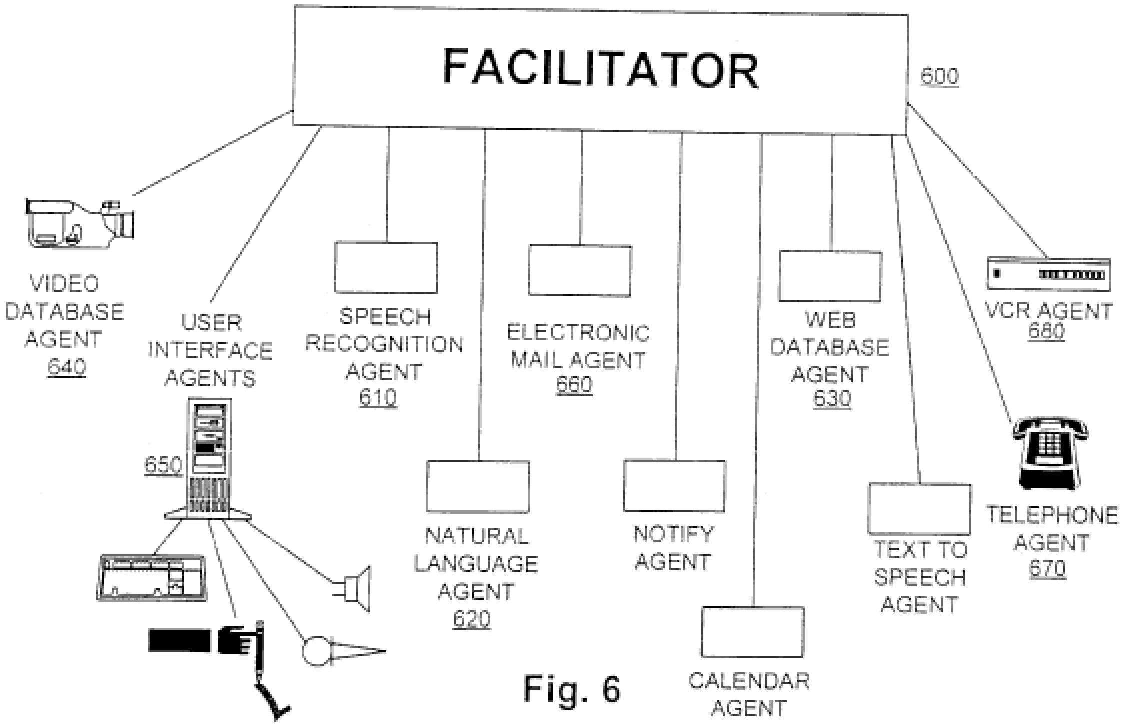
### A. The ’718 Patent

The ’718 patent issued from Application No. 09/608,872 (“the ’872 application”), filed on June 30, 2000, and claims a March 17, 1999 priority date. (*See* Ex. 1001, Cover; *see also* Exs. 1005, 1007, 1009-1011.) The ’718 patent “relates generally to the navigation of electronic data by means of spoken natural language requests, and to feedback mechanisms and methods for resolving the errors and ambiguities that may be associated with such requests.” (Ex. 1001, 1:22-26; *see also* Ex. 1002, ¶¶38-40.)

The ’718 patent uses the then-existing Open Agent Architecture (OAA). (Ex. 1001, 3:46-48, 13:16-19, 14:27-29, FIG. 6 (reproduced below); Ex. 1002, ¶41.) The OAA includes multiple “autonomous entities, or agents” and a facilitator agent. (Ex. 1007, 4:20-21; Ex. 1001, FIG. 6 (reproduced below); Ex. 1002, ¶41.)

---

[1] Petitioner submits the declaration of Dr. Dan R. Olsen Jr. (Ex. 1002), an expert in the field of the ’718 patent. (Ex. 1002, ¶¶1-9; Ex. 1003.)



(Ex. 1001, FIG. 6.)

"[A]n agent registers with its parent facilitator a specification of the capabilities and services it can provide," and "[w]hen a facilitator determines that the registered capabilities of one of its client agents will help satisfy a current goal or sub-goal thereof, the facilitator delegates that sub-goal to the client agent … ." (*Id.*, 13:36-45; *see also id.*, 1:5-18, 13:19-22, 13:34-51; Ex. 1007, 6:10-13; Ex. 1002, ¶42.)

### B. Prosecution History of the '718 Patent

During prosecution, in response to anticipation rejections issued by the Examiner (Ex. 1004, 138-47), the Applicants amended each then-pending independent claim to add the limitation "wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television." (*Id.*, 150-58.) After the Examiner issued another Office Action containing obviousness rejections (*id.*, 178-81), the Applicants presented arguments regarding the limitation "a portable remote control device or a set-top box for a television" without further amending the claims. (*Id.*, 185-86; *see also id.*, 183-186.) The Examiner then allowed the claims. (*Id.*, 193-95.)

### C. The Prior Art

*Cheyer*, whose authors are two of the named inventors of the '718 patent, describes "how multiple input modalities may be combined to produce more natural user interfaces." (Ex. 1012, 1.) *Cheyer*'s multimodal application uses the then-existing Open Agent Architecture to implement "a distributed network of heterogeneous software agents" for distributed processing regarding various tasks. (*Id.*; Ex. 1002, ¶47.)

*Cheyer* discloses various examples of receiving a spoken natural language (e.g., English) request for desired information from a user on a PC or a handheld PDA. (Ex. 1012, 4-6, 11; Ex. 1002, ¶48.) The spoken English request is processed by a speech recognition (SR) agent and a natural language (NL) parser agent to recognize a speech string in the user's speech input and translate the recognized request into a format called Interagent Communication Language that software agents can handle. (Ex. 1012, 7, 9-11; Ex. 1002, ¶49.) The SR and NL agents are among several agents (shown below in Figure 3 of *Cheyer*) that are implemented using the Open Agent Architecture to perform various tasks to service the user's request. (Ex. 1012, 7-12; Ex. 1002, ¶49.)

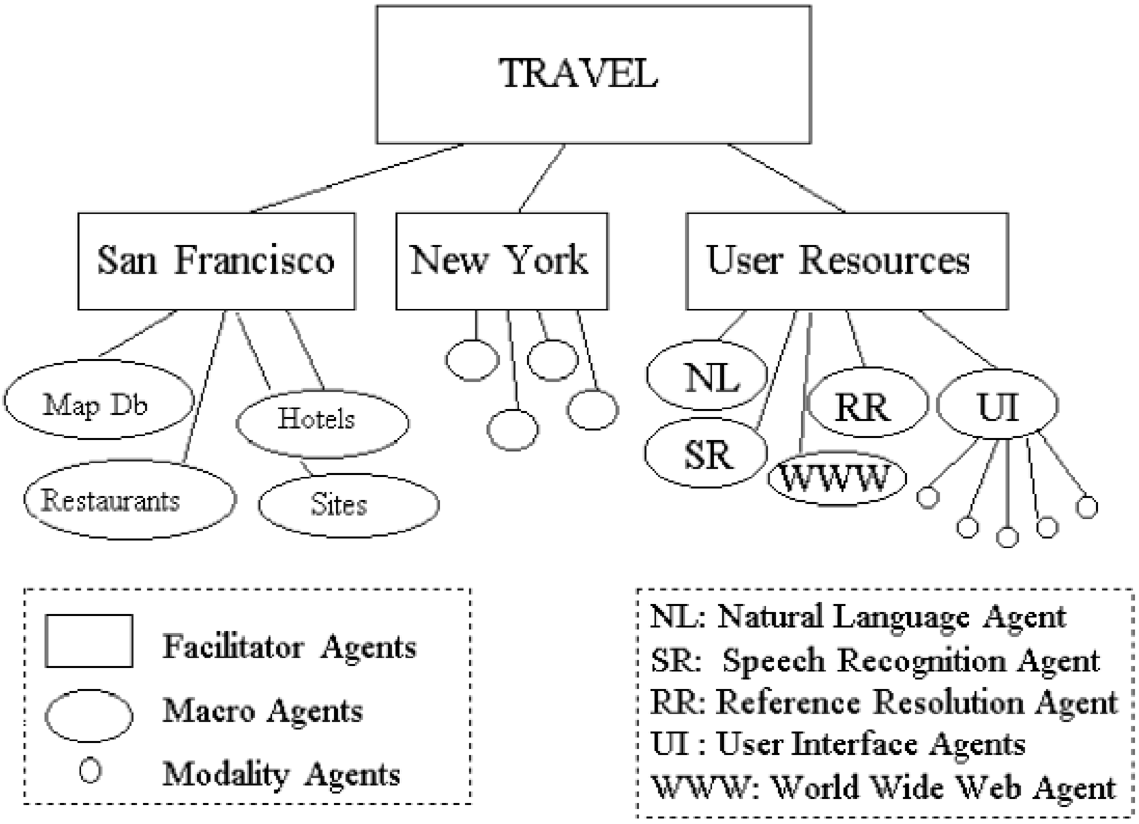


Figure 3: Agent Architecture for Map Application

(Ex. 1012, FIG. 3.) *Cheyer* discloses that "[t]he architecture for the OAA … uses a hierarchical configuration where client agents connect to a 'facilitator' server," also referred to as a "facilitator agent." (*Id.*, 7, 9.) *Cheyer* discloses that the facilitator agent "records the published functionality of [its] sub-agents." (*Id.*, 8; *see also* Ex. 1002, ¶¶50-51.)

*Shwartz*, *Thrift*, *Dureau*, *Johnson*, and *Simmers* provide additional details on many of the well-known user interface and networking technologies described in the '718 patent. (Ex. 1002, ¶¶52-61; *see also id.*, ¶¶16-37 (discussing the state of the art).)

## VIII. CLAIM CONSTRUCTION

The '718 patent will expire on January 5, 2019, which is during the likely pendency of this IPR proceeding should the Board institute review. Accordingly, the claims should be construed under the standard set forth in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc). Under *Phillips*, claim terms are given their ordinary and customary meanings, as would have been understood by a POSITA, at the time of the invention, having taken into consideration the language of the claims, the specification, and the prosecution history of record. *See, e.g.*, *Cisco Sys., Inc., v. AIP Acquisition, LLC*, IPR2014-00247, Paper No. 20 at 2-3 (July 10, 2014). The Board, however, only construes the claims when necessary to resolve the underlying controversy. *Toyota Motor Corp. v. Cellport Systems, Inc.*, IPR2015-00633, Paper No. 11 at 16 (Aug. 14, 2015) (citing *Vivid Techs., Inc. v.*

*Am. Sci. & Eng'g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999)). Petitioner provides below the construction of various terms that are relevant to this proceeding.[2]

### A. "Navigation Query"

Claims 1, 4, 10, 13, 19, and 22 recite "navigation query." In district court, Patent Owner has argued that "navigation query" should be construed as "an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information." (Ex. 1019, 2.) This construction corresponds to the indication in the specification that "[a] 'navigation query' means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information." (Ex. 1001, 8:65-9:1.) For purposes of this Petition, Petitioner applies Patent Owner's proposed construction of "navigation query." (Ex. 1002, ¶44.)

---

[2] Petitioner reserves all rights to raise claim construction and other arguments in district court. For example, Petitioner has not necessarily raised all challenges to the '718 patent, including those under 35 U.S.C. § 112, given the limitations placed by the Rules governing this proceeding.

### B. “Code Segment [That]” and “Logic[,] Operable To”

Claims 10 and 13 recite “code segment[s] [that] [perform various functions],” and claims 19 and 22 recite various “logic” “operable to [perform various functions].” Petitioner identifies below for each of the foregoing claim terms the identified function (in **bold**) and corresponding structure that performs such identified function, under the assumption that these terms invoke 35 U.S.C. § 112 ¶ 6. 37 C.F.R. § 42.104(b)(3).

| Claim Term and **Identified Function** | Corresponding Structure |
|---|---|
| “code segment that **receives a spoken request for desired information from the user utilizing the mobile information appliance of the user**” (claim 10) | As explained below, each of these elements in the left column recites function without sufficient structure for performing the function. However, for purposes of this proceeding, the structure should be software running on a microprocessor configured to perform the identified functions or |
| “code segment that **renders an interpretation of the spoken request**” (claim 10) | |
| “code segment that **constructs a navigation query based upon the interpretation**” (claim 10) | |
| “code segment that **utilizes the navigation query to select a portion of the electronic data source**” (claim 10) | |

| Claim Term and **Identified Function** | Corresponding Structure |
|---|---|
| "code segment that **transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user**" (claim 10)[3] | equivalents thereof. |
| "code segment that **solicits additional input from the user, including user interaction in a modality different than the original request**" (claim 13) | |
| "code segment that **refines the navigation query, based upon the additional input**" (claim 13) | |

---

[3] Claim 19 recites a similar limitation: "(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user." Because of the "infrastructure" recited in the claim and the corresponding disclosure in the specification of the '718 patent (Ex. 1001, 4:48-55), Petitioner believes this limitation of claim 19 does not invoke 35 U.S.C. § 112 ¶ 6, but to the extent the Board decides otherwise, this limitation of claim 19 should be construed in the same manner as the corresponding limitation of claim 10.

| Claim Term and **Identified Function** | Corresponding Structure |
|---|---|
| "code segment that **uses the refined navigation query to select a portion of the electronic data source**" (claim 13) | |
| "spoken language processing logic, operable to **render an interpretation of the spoken request**" (claim 19) | |
| "query construction logic, operable to **construct a navigation query based upon the interpretation**" (claim 19) | |
| "navigation logic, operable to **select a portion of the electronic data source using the navigation query**" (claim 19) | |
| "user interaction logic operable to **solicit additional input from the user, including user interaction in a modality different than the original request**" (claim 22) | |
| "query refining logic operable to **refine the navigation query based upon the additional** | |

| Claim Term and **Identified Function** | Corresponding Structure |
|---|---|
| **input**" (claim 22) | |

A structure disclosed in the specification qualifies as corresponding structure only if it is clearly linked by the patent's specification (or possibly the prosecution history) to performing the claimed function. *See Default Proof Credit Card Sys., Inc. v. Home Depot U.S.A., Inc.*, 412 F.3d 1291, 1298 (Fed. Cir. 2005); *Gracenote, Inc. v. Iceberg Indus., LLC*, IPR2013-00551, Paper No. 6 at 15 (Feb. 28, 2014). Where a means-plus-function term is directed to software, the specification must "disclose an algorithm for performing the claimed function." *Williamson v. Citrix Online, LLC*, 792 F.3d 1339, 1352 (Fed. Cir. 2015).

For the terms in the table above, the only corresponding structure under 35 U.S.C. § 112 ¶ 6 disclosed is software running on a processor. For example, the specification of the '718 patent discloses "a general-purpose hardware microprocessor" for implementing various embodiments. (Ex. 1001, 6:65-7:3.)

The '718 patent specification does not describe the "code segment[s]" and "logic[s]" as claimed in claims 10, 13, 15, 19, and 22, other than by way of functional description. Given that the "code segment[s]" and "logic[s]" refer to computer software, and given that none of the identified functions is a "generic function," the corresponding structure for such terms requires an algorithm.

*Williamson*, 792 F.3d at 1352. However, beyond repeating some claim language for some identified functions, the ’718 patent does not disclose an algorithm that corresponds to the identified functions of these terms. Thus, with respect to each of the identified functions for these terms discussed above, the ’718 patent simply discloses a “black box” that performs some function, “[b]ut how it does so is left undisclosed.” *Blackboard, Inc. v. Desire2Learn, Inc.*, 574 F.3d 1371, 1383 (Fed. Cir. 2009).

For purposes of this proceeding, as required by 37 C.F.R. § 42.104(b)(3), Petitioner submits that the corresponding structure for each of the above-identified functions of the terms listed above should be software running on a microprocessor configured to perform the identified functions or equivalents thereof under 35 U.S.C. § 112 ¶ 6.[4] (Ex. 1002, ¶¶45-46.)

## IX. DETAILED EXPLANATION OF GROUNDS

As discussed below, the challenged claims are unpatentable in view of the prior art.

---

[4] Petitioner does not concede that claims 10 and 19 and their dependent claims are not indefinite. Moreover, the analysis below addresses these claims even if the terms do not invoke 35 U.S.C. § 112 ¶ 6.

## A. Ground 1: *Cheyer*, *Shwartz*, and *Thrift* Render Obvious Claims 1-4, 6, 8-10, 12, 13, 15, 17-19, 21, 22, 24, 26, 27

### 1. Claim 1

#### i) "A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:"

To the extent the preamble of claim 1 is limiting, *Cheyer* discloses the limitations therein. (Ex. 1002, ¶63.) For instance, *Cheyer* discloses a method for processing input provided by a user via "spoken natural language" (Ex. 1012, 4) ("speech-based") to enable the user "to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web" (*id.*) ("navigation of an electronic data source"). (Ex. 1002, ¶64; *see also* Ex. 1012, 11-12 (providing an example where a user's speech-based query is processed to provide the user with requested information).)

More specifically as to "*speech-based*," *Cheyer* discloses an "application [that] is distinguished by a synergistic combination of handwriting, gesture and *speech* modalities." (Ex. 1012, 1 (emphasis added).) In particular, *Cheyer* provides the user with the ability to enter natural language input via a variety of modalities, including speech-based, and explains benefits associated with such a speech-based method. (*Id.*, 2-3; Ex. 1002, ¶65.) *Cheyer* provides various examples of spoken input requests by a user. (*See, e.g.*, *id.*, 5-6, 11; Ex. 1002, ¶¶65-66.)

More specifically as to "*navigation of an electronic data source*," *Cheyer* discloses navigation of data sources such as remote databases on the World Wide Web. (Ex. 1012, Abstract, 6, 7, 10, 12; Ex. 1002, ¶67.) *Cheyer* discloses that the remote database is located at one or more network servers located remotely from a user. (Ex. 1002, ¶68.) For example, *Cheyer* discloses "access to existing data sources including the World Wide Web" (Ex. 1012, Abstract), and explains that its system enables "a mobile system that provides a synergistic pen/voice interface to remote databases" (*id.*, 6). A POSITA would have understood that the way a user's device retrieved information from the World Wide Web was by contacting a remote server (e.g., web server) that could transmit the information to the user's device. (Ex. 1002, ¶68.) Indeed, the existence of servers on a network that enabled a user to access data remotely was one of the fundamental principles of the World Wide Web. (*Id.*)

A POSITA would have understood that *Cheyer* necessarily discloses that a data link is established between the user's mobile device ("mobile information appliance of the user") and the remote server ("one or more network servers"). A "handheld PDA" (Ex. 1012, 4, 6) with a "mobile handheld interface" (*id.*, Abstract) as disclosed by *Cheyer* is a "mobile information appliance of the user" as recited in the preamble. (Ex. 1002, ¶69.) *Cheyer* discloses that the "mobile system [] provides [an] interface to remote databases," and thus discloses that the user's mobile device communicates with the remote databases. (Ex. 1012, 6; Ex. 1002, ¶69; *see also* Ex. 1012, Abstract ("access to existing data sources including the World Wide Web; and a mobile handheld interface"), 4 ("Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web"), 7 ("access to various heterogeneous data and knowledge sources"), 12 ("mobile … interface providing … access to heterogeneous distributed knowledge sources").) Such communication reflects a data link between the user's mobile device and the remote server. (Ex. 1012, 6; Ex. 1002, ¶69.)

(*See also infra* Sections IX.A.1.ii-vi regarding the remaining limitations of this claim.)

#### ii) [1.a] "(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;"

*Cheyer* in combination with *Thrift* discloses this limitation. (Ex. 1002, ¶71.) For instance, *Cheyer* discloses various examples of receiving a spoken request for desired information from a user. (Ex. 1012, 5 ("'How far is the hotel from Fisherman's Wharf?'" and "'Show me all available information about Alcatraz'"), 11 ("A user speaks: 'How far is the restaurant from this hotel?'"); *see also id.*, 5-6; Ex. 1002, ¶71.) In each of these examples, the user is requesting desired information via a spoken request. (Ex. 1002, ¶72.)

*Cheyer* discloses that the user's mobile computing device receives the spoken request from the user utilizing the user's mobile computing device ("mobile information appliance of the user"). (Ex. 1012, 4 ("[T]he system permits the user to [provide] spoken natural language …. The user interface must be light and fast enough to run on a handheld PDA"), 6 ("The user interface runs on pen-equipped PC's or a Dauphin handheld PDA … using either a microphone or a telephone for voice input."), Abstract ("The application is distinguished by … a mobile handheld interface"); Ex. 1002, ¶73.) *Cheyer* also discloses that a micro agent associated with a speech recognition agent receives the spoken request after it is received by the user's computing device. (*See* Ex. 1012, 9, 11; Ex. 1002, ¶74.)

*Cheyer* discloses that the device that receives voice input from the user is a portable device. (Ex. 1012, Abstract ("mobile handheld interface"), 4 ("handheld PDA"), 6 ("mobile system"), 12 ("mobile … interface"); Ex. 1002, ¶75.) *Cheyer* further discloses that the user's mobile device communicates with a remote server to cause the remote server to retrieve information responsive to a user's query (e.g., "Show me all available information about Alcatraz") and send such retrieved information to the user's device, e.g., so that the user can see all available information about Alcatraz. (Ex. 1012, 5; Ex. 1002, ¶75; *see also* Ex. 1012, 4 ("Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web"), 6 ("mobile system that provides [an] interface to remote databases"), Abstract ("access to existing data sources including the World Wide Web; and a mobile handheld interface"); *supra* Section IX.A.1.i (citations and analysis regarding data link and network server located remotely from a user); *infra* Sections IX.A.1.v-vi.) Because *Cheyer*'s mobile device of the user remotely causes a server to take prescribed actions (e.g., retrieve requested information and send it to the mobile device), the mobile device is a remote control device. (Ex. 1002, ¶75.) *Cheyer* further discloses that the user's mobile device can be a PDA (Ex. 1012, 4, 6), and thus discloses a *portable* remote control device. (Ex. 1002, ¶75.)

To the extent_*Cheyer* does not expressly disclose that "said mobile information appliance comprises a … remote control device or a set-top box *for a television*" as recited in limitation [1.a], it would have been obvious in view of *Thrift* to modify *Cheyer*'s process to include such features. (Ex. 1002, ¶76.)

*Thrift* "relates generally to voice recognition devices" and discloses examples of voice-activated devices for controlling a processor-based host system. (Ex. 1015, 1:9-10; Ex. 1002, ¶77; *see also* Ex. 1015, Abstract, 2:42, 2:43-46.)

Thus, *Thrift* is in the same technical field as *Cheyer* (e.g., voice interface for retrieving information desired by a user). (Ex. 1012, Abstract; Ex. 1002, ¶77.)

A POSITA implementing *Cheyer*'s process and system would have had reason to consider the teachings of *Thrift* for enhancing the feature set and functionality of *Cheyer*'s process and system. (Ex. 1002, ¶78.) *Thrift* describes a system that "makes information on the Web more accessible and useful" and explains that "[s]peech control brings added flexibility and power to the Web interface and makes access to information more natural," and a POSITA would have recognized those attributes as being pertinent to *Cheyer*'s process, which similarly involves a voice interface for retrieving information from the Web. (Ex. 1015, 2:15-18; Ex. 1002, ¶78; *see also supra* Section IX.A.1.i (citations and analysis regarding *Cheyer*'s voice interface for retrieving information from the Web).)

Additionally, a POSITA would have found *Thrift*'s disclosure of a system that interprets a user's command such as "What's on TV tonight" or "Give me the weather" to be similar to *Cheyer*'s disclosure of a system that provides information to the user based on spoken commands. (Ex. 1015, 3:60, 4:58; Ex. 1012, Abstract, 4-6, 9-11; *see also* Ex. 1015, 4:25-26, 4:41-42, 4:57-58; Ex. 1002, ¶79.)

Having looked to *Thrift*, a POSITA would have seen that *Thrift* discloses a wireless "voice-activated remote control device." (Ex. 1015, 2:39-40; Ex. 1002,

¶80; *see also* Ex. 1015, 1:66-67, 2:37-39.) *Thrift* further discloses a remote control device in the context of controlling a television. (Ex. 1015, 2:43-46 ("voice-controlled device for controlling … a television").)

A POSITA would have been motivated in light of the teachings of *Thrift* to configure *Cheyer*'s process and system so that the handheld device that receives input from the user ("said mobile information appliance") comprises a portable remote control device for a television. (Ex. 1002, ¶81.) For example, a POSITA would have recognized that just like *Cheyer*'s handheld PDA which receives speech input, *Thrift*'s voice-activated control unit 10 is wireless and includes a processor, memory, display, and microphone to receive voice input. (Ex. 1015, 2:37, 3:10-11, 3:11-12, 2:59-62, Abstract, FIG. 1 (reproduced below); Ex. 1002, ¶81.)

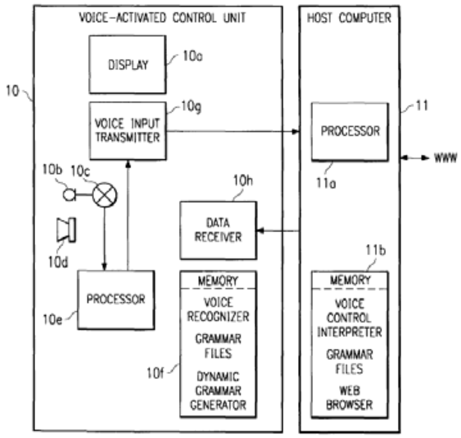


*FIG. 1*

(Ex. 1015, FIG. 1.)

A POSITA would further have recognized the benefits of implementing the device used in *Cheyer*'s process to be a remote control device for a television. (Ex. 1002, ¶82.) For example, a POSITA would have recognized that configuring the device to be a portable remote control device for a television would have enabled the user to retrieve information via a broader set of devices, e.g., via a television as disclosed in *Thrift*. (Ex. 1015, 2:44-46; Ex. 1002, ¶82.)

A POSITA would further have recognized that configuring a device to be a remote control device for a television would have been a familiar, user-friendly configuration because remote controls for televisions were well-known long before the alleged invention of the '718 patent. (Ex. 1002, ¶83.) Implementing such a configuration would have been straightforward, because *Thrift*'s control unit 10 includes a wireless transmitter 10g and receiver 10h for remotely controlling and communicating with another device and a POSITA would have known how to program *Cheyer*'s handheld PDA, which similarly includes wireless communication components, to be a remote control for a television. (*Id.*)

Furthermore, a POSITA would have recognized that configuring *Cheyer*'s mobile device to be a portable remote control device for a television would have been a predictable implementation, because it was well known at the time of the alleged invention of the '718 patent to provide voice input to components for a television. (Ex. 1002, ¶84.) For example, *Dureau*[5] discloses a system in which a "user can use a microphone or a telephone handset to provide voice data to the system," whereby the "microphone may be connected to [a] set-top box, or it may be built into a remote control for the system," and thereafter the "voice data is transmitted to the server, which uses voice recognition software to convert the voice data into textual data." (Ex. 1016, 10:56-67; Ex. 1002, ¶84.)

The above configuration would have been a mere combination of known components and technologies (e.g., *Cheyer*'s functionality relating to a voice interface for a device that remotely controls another device, and *Thrift*'s disclosure of a voice-controlled remote control device for a television), according to known methods (e.g., a POSITA knew how to program a device to implement wireless

---

[5] *Dureau* is only cited for claim 1 to demonstrate knowledge of a POSITA and is not relied upon as a reference in this unpatentability ground.

communication to remotely control a television), to obtain predictable results (e.g., a voice-controlled remote control device for a television that could be used to provide desired information to a user). (Ex. 1002, ¶85.) *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 416 (2007).

#### iii) [1.b] "(b) rendering an interpretation of the spoken request;"

*Cheyer* discloses this limitation. (Ex. 1002, ¶86.) For instance, *Cheyer* discloses that a speech recognition agent recognizes a spoken English request and a "Natural Language (NL) Parser Agent" translates the request into the Interagent Communication Language (ICL). (Ex. 1012, 7, 9-11, FIG. 3 (reproduced below); Ex. 1002, ¶86.)

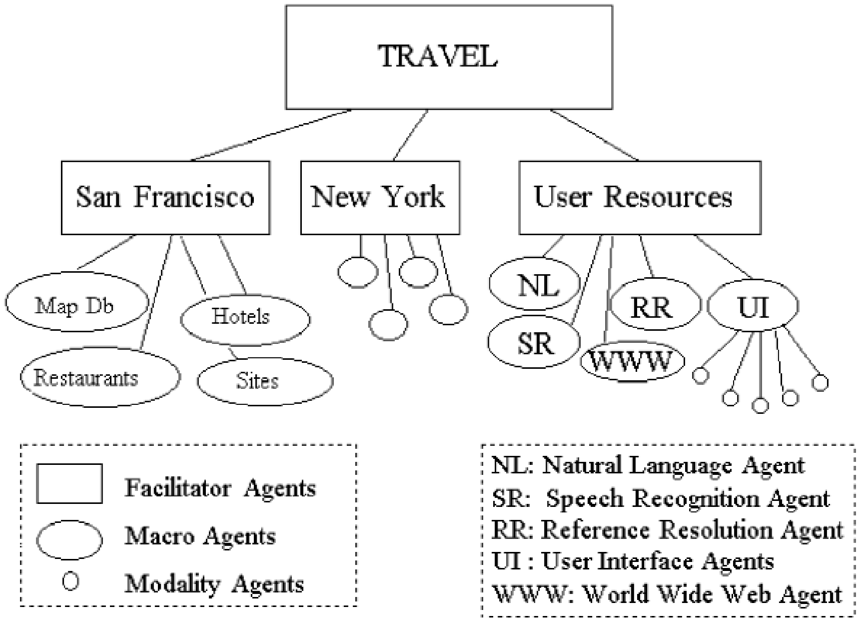


Figure 3: Agent Architecture for Map Application

(Ex. 1012, FIG. 3 (showing speech recognition agent and NL agent).)

The speech recognition and ICL translation of the user's speech input constitute an "interpretation of the spoken request," so *Cheyer* discloses that the speech recognition agent and NL parser agent "render[] an interpretation of the spoken request." (Ex. 1002, ¶87.) In fact, the '718 patent specification discloses the same use of a speech recognition agent and NL parser agent as disclosed in *Cheyer*. (*See, e.g.*, Ex. 1001, 14:33-36 (explaining that a "speech recognition agent 610" and "natural language (NL) agent 620" render an "interpretation in ICL format"); Ex. 1012, 7, 9-11; Ex. 1002, ¶87.)

### iv) [1.c] "(c) constructing a navigation query based upon the interpretation;"

*Cheyer* in combination with *Shwartz* discloses this limitation. (Ex. 1002, ¶88.) For instance, *Cheyer* discloses that based on the interpretation provided by the speech recognition agent and NL parser agent, a domain agent "sends database requests" asking for information related to the user's request, e.g., coordinates of items such as a reference or hotel. (Ex. 1012, 12; Ex. 1002, ¶88.)

Therefore, *Cheyer* discloses a "navigation query" because *Cheyer*'s domain agent sends a database request ("navigation query") that enables the desired information to be retrieved for the user. (*Supra* Section VIII.A; Ex. 1002, ¶89.) *Cheyer*'s database request is a navigation query because it is an electronic query structured appropriately so as to navigate a data source of interest in search of desired information. (*Supra* Section VIII.A; Ex. 1002, ¶89.)

While *Cheyer* may not expressly describe the details of "constructing a navigation query based upon the interpretation," it would have been obvious in view of *Shwartz* to implement such features in *Cheyer*'s process. (Ex. 1002, ¶90.) For example, while *Cheyer* discloses using database requests to retrieve information from a database to service a user's request (Ex. 1012, 11, 12; *see also id.*, 5, 6), *Cheyer* does not provide details regarding constructing such database requests, but *Shwartz* discloses constructing a database query to navigate a database in search of desired information, as set forth below. (Ex. 1002, ¶90.)

*Shwartz*, which is in the same technical field as *Cheyer* (e.g., natural language interface for servicing a user's request), discloses a "database retrieval system having a natural language interface" and further discloses that "[a] *database query is generated* … , enabling the retrieval and aggregation of data from [a] database to satisfy [a] natural language query." (Ex. 1013, Abstract (emphasis added); Ex. 1002, ¶91.) For example, *Shwartz* discloses "retrieval of information from the application database in response to a query represented by the meaning representation." (Ex. 1013, 9:25-27.)

*Shwartz* explains that "[a] navigator and query language generator 38 is used to define optimal navigation paths through the database tables and columns to respond to the query, and to generate a meta-query language ('MQL')," and "[t]he metaquery language is used by a reporter and database access system 40 to *generate the code (e.g., structured query language ('SQL') code)* to actually retrieve the information from the application database." (*Id.*, 9:28-35 (emphasis added); *see also id.*, 7:19-22, 17:1-19 (disclosing details regarding how to locate information from application database 32 responsive to a query); Ex. 1002, ¶92.)

Thus, *Shwartz* teaches details of constructing a query suitable for retrieving, from a database (such as *Cheyer*'s remote databases), information desired by a user. (Ex. 1002, ¶93.) A POSITA would have understood *Shwartz* to teach constructing a "*navigation* query" because *Shwartz*'s foregoing generated query (e.g., SQL query) is an electronic query structured appropriately so as to navigate a particular data source in search of desired information. (*Supra* Section VIII.A; Ex. 1002, ¶93.)

Because *Cheyer*'s database request "ask[s] for the coordinates of the items in question" (e.g., the coordinates of the restaurant and the hotel referenced by the user's input query "How far is the restaurant from this hotel?") and the items in question are contained in the user's input query that is processed by the speech recognition agent and NL parser agent to interpret the meaning of the words in the input query, a POSITA would have been motivated to configure the combined *Cheyer*-*Shwartz* process to construct the database query based upon the interpretation that is rendered, similarly to the arrangement in *Shwartz*. (Ex. 1012, 11-12; Ex. 1013, 7:56-60, 7:54-55; *see also id.*, 9:20-35, FIG. 1; Ex. 1002, ¶94.)

In other words, a POSITA would have been motivated to construct the database query in the combined process based upon the interpretation of the user's spoken request so that the database query could properly specify information to be retrieved from *Cheyer*'s remote database. (Ex. 1002, ¶95.)

A POSITA would have been capable of implementing the above configuration and would have had a reasonable expectation of success regarding the outcome. (Ex. 1013, Abstract; Ex. 1002, ¶96.) This would have been a straightforward implementation that merely involved constructing a navigation query to access a database in a predictable manner. (Ex. 1002, ¶96.) Indeed, such an implementation would have been a mere combination of elements and technologies (e.g., a database request for servicing a query, as taught by *Cheyer*, and construction of a database query, i.e., database request, as taught by *Shwartz*), according to known methods (e.g., *Shwartz* describes how to construct the query, and *Cheyer* describes its role in a system for servicing a user's request), to provide predictable results (e.g., retrieving information desired by the user from a database). (*Id.*) *KSR*, 550 U.S. at 416.

#### v) [1.d] "(d) utilizing the navigation query to select a portion of the electronic data source; and"

*Cheyer* alone and/or in combination with *Shwartz*[6] discloses this limitation. (Ex. 1002, ¶97.) For instance, *Cheyer* discloses that a database agent utilizes the navigation query to retrieve from a database information requested by a user ("select a portion of the electronic data source"). (*Id.*, ¶97.) *Cheyer* discloses various examples of such "portion[s] of the electronic data source," such as "maps

---

[6] As discussed above for limitation [1.c], it would have been obvious in view of *Shwartz* to modify *Cheyer*'s process to construct a "navigation query." It would also have been obvious to configure the combined *Cheyer*-*Shwartz* process to implement the features relating to "navigation query" in limitation [1.d] and claim 4. (Ex. 1002, ¶97 n.4.)

for each city, as well as icons, vocabulary and information about available hotels, restaurants, movies, theaters, municipal buildings and tourist attractions" (Ex. 1012, 10), "the French restaurants within 1 mile of this hotel" (*id.*, 5) or "all available information about Alcatraz" (*id.*, 5).

*Cheyer* discloses that a type of agent called a "facilitator" routes information to agents in the Open Agent Architecture. (*See id.*, 7 ("Facilitators provide content-based message routing, global data management, and process coordination for their set of connected agents."), 8 ("when queries arrive in Interagent Communication Language form, [facilitators] are responsible for breaking apart any complex queries and for distributing goals to the appropriate agents"); *see also id.*, 9 ("facilitator agent"); Ex. 1002, ¶98.)

*Cheyer* discloses that database agents provide information (e.g., about maps, places of interest, movies, etc.) relevant to the user's request. (Ex. 1012, 10 ("database agents provide maps for each city, as well as icons, vocabulary and information about available hotels …"); Ex. 1002, ¶99.)

*Cheyer*'s database agents retrieve information from a database based on database requests. (Ex. 1012, 10 ("a domain agent will try to resolve the definite reference by sending database agent requests"); Ex. 1002, ¶100.) Thus, when a database request is constructed for retrieving information from a database in response to a user's input such as "Display the French restaurants within 1 mile of this hotel" (Ex. 1012, 5) or "Show me all available information about Alcatraz" (*id.*), a corresponding database request is routed to a database agent that services the request by utilizing the database request ("navigation query") to access the database. (Ex. 1002, ¶100.)

While *Cheyer* discloses "access to existing data sources" (Ex. 1012, 1), "access to various heterogeneous data and knowledge sources" (*id.*, 7), "access [to] a wide variety of data sources, including information stored in HTML form on the World Wide Web" (*id.*, 4), and various types of databases, including "Prolog databases, X-500 hierarchical databases, and data loaded automatically by scanning HTML pages from the World Wide Web (WWW)" (*id.*, 10), *Cheyer* does not expressly disclose that the database agent "select[s] a portion" of the disclosed electronic data source. However, a POSITA would have understood that *Cheyer* necessarily discloses that feature. (Ex. 1002, ¶101.) A POSITA would have had this understanding because "database requests" (Ex. 1012, 12) were well known to be for retrieving or selecting a portion of a database. (Ex. 1002, ¶101.) If a portion of the database that contains the "maps for each city" or "information about available hotels … and tourist attractions" (Ex. 1012, 10) were not selected by *Cheyer*'s database agent, then the database agent would not have been able to provide the information that the user requested. (Ex. 1002, ¶101.)

To the extent *Cheyer* does not disclose "select[ing] a portion of the electronic data source," it would have been obvious in view of *Cheyer* and *Shwartz* to implement this feature in *Cheyer*'s process. (*Id.*, ¶102.) *Shwartz* discloses "retrieval and aggregation of data from [a] database to satisfy [a] natural language query" (Ex. 1013, Abstract) and "identify[ing] an optimal set of database elements to satisfy the query" (*id.*, 17:10-11), e.g., by choosing particular "tables and columns" (*id.*, 9:24-27). Additionally, *Shwartz* discloses "generat[ing] … code (e.g., structured query language ('SQL') code) to actually retrieve the information from the application database" (*id.*, 9:33-35), and a POSITA would have understood that SQL code (e.g., a SELECT statement in SQL code) was intended to select a portion of a database. (Ex. 1002, ¶102; *see also* Ex. 1013, 7:19-22.)

A POSITA would have been motivated, in light of the teachings of *Cheyer* and *Shwartz*, to configure *Cheyer*'s process to select a portion of any of the databases disclosed by *Cheyer*. (Ex. 1002, ¶103.) A POSITA would have recognized that selecting a portion of a database responsive to the user's request would have enabled the combined *Cheyer-Shwartz* process and system to provide desired information to the user. (*Id.*) This would have been a straightforward configuration, because it would have been merely a combination of known components and technologies (e.g., *Cheyer*'s database and database requests, and *Shwartz*'s "structured query language ('SQL') or other code" for retrieving a portion of a database (Ex. 1013, 7:19-22)), according to known methods (e.g., retrieving information from a database using database requests), to obtain predictable results (selecting a portion of a database in response to a database request). (Ex. 1002, ¶¶103-104.) *KSR*, 550 U.S. at 416.

#### vi) [1.e] "(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user."

*Cheyer* alone and/or in combination with *Shwartz* discloses this limitation. (Ex. 1002, ¶105.) For instance, in the examples of "Display the French restaurants within 1 mile of this hotel" (Ex. 1012, 5) or "Show me all available information about Alcatraz" (*id.*), *Cheyer* discloses displaying the French restaurants within 1 mile of the hotel specified by the user or displaying all available information about Alcatraz. (Ex. 1002, ¶105; *see also* Ex. 1012, 10, 12.) *Cheyer* also discloses "access to existing data sources including the World Wide Web." (Ex. 1012, Abstract.)

Based on the foregoing disclosures, a POSITA would have understood that *Cheyer* necessarily discloses transmitting the selected portion of the electronic data source from the remote server at which such data sources are located ("the network server") to the user's mobile computing device ("the mobile information appliance of the user"). (Ex. 1002, ¶106.) For example, if such data were not transmitted from the remote server to the user's mobile device, the user could not have obtained the information that he/she desired. (*Id.*) Indeed, *transmitting* data from a remote server to a user's computing device was known to be a necessary aspect of data communications involving the Web, which *Cheyer* discloses. (Ex. 1012, Abstract; Ex. 1002, ¶106.)

To the extent *Cheyer* does not disclose transmitting the selected portion of the electronic data source from the remote server to the user's mobile device, it would have been obvious in view of *Shwartz* to implement such features. (Ex. 1002, ¶107.) As discussed above for limitation [1.d], it would have been obvious in view of *Shwartz* to select a portion of the electronic data source (*supra* Section IX.A.1.v), and in view of *Shwartz*'s disclosure of displaying retrieved data on a user's computer it would further have been obvious to configure the combined process to transmit the selected portion from the remote network server to the user's mobile device. (Ex. 1013, 5:9-11; Ex. 1002, ¶107.) A POSITA would have known how to implement data communications involving the Web, which *Cheyer* discloses (Ex. 1012, Abstract), and would have been motivated to implement such transmitting in order to achieve a working application as disclosed in *Cheyer*. (*Id.*) Indeed, a POSITA would not only have been motivated but would have naturally expected to configure *Cheyer*'s process to transmit the selected portion of the electronic data source from the remote server to the user's mobile device, in order to achieve *Cheyer*'s objective of enabling a user "to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web." (Ex. 1012, 4; Ex. 1002, ¶¶107-108.)

This would have been a mere combination of known components and technologies (e.g., *Cheyer*'s disclosure of an application that retrieves information from a remote data source such as one located on the Web, *Cheyer*'s disclosure of a PDA that a POSITA would have known was capable of receiving information transmitted by a remote server, and *Shwartz*'s disclosure of displaying retrieved data on a user's computer), according to known methods (e.g., implementing data communications involving the Web in a known manner), to obtain predictable results (e.g., sending information from a remote server to the user's mobile device). (*Id.*, ¶109.) *KSR*, 550 U.S. at 416.

### 2. Claims 2 and 3

#### i) [2.a]/[3.a] "The method of claim 1, wherein the step of rendering the interpretation of the spoken request is performed by the mobile information appliance."[7]

While *Cheyer* discloses a "server machine which will manage … natural language processing and speech recognition for the application" (Ex. 1012, 6), it would have been obvious in view of *Thrift* to configure the combined *Cheyer-Thrift-Shwartz* process to perform the speech recognition and natural language processing ("the step of rendering the interpretation of the spoken request") at the user's mobile computing device ("the mobile information appliance"). (Ex. 1002, ¶110.)

*Cheyer* discloses "[t]he user interface must be light and fast enough to run on a handheld PDA while able to access applications and data that *may* require a more powerful machine" (Ex. 1012, 4), which suggests that in some situations

---

[7] Claim 2 appears to have issued with a printing error. Specifically, during prosecution claim 2 recited "at the one or more network servers" instead of "by the mobile information appliance." (Ex. 1004, 67.) Nonetheless, Petitioner addresses claim 2 as issued here. To the extent claim 2 is interpreted to require "at the one or more network servers," instead of "by the mobile information appliance," that interpretation is addressed below in Section IX.B.

(e.g., when the user's handheld PDA is sufficiently powerful) a more powerful machine (e.g., server remote from the PDA) may not be needed. (Ex. 1002, ¶111.) A POSITA would have understood *Cheyer*'s foregoing disclosure as providing guidance as to when a remote server for performing speech recognition and natural language processing would or would not be appropriate (i.e., the resource capabilities of the PDA are central to this issue). (*Id.*)

*Thrift*, in the same technical field as *Cheyer* (e.g., providing information to a user based on voice input), discloses a client-server architecture in the speech processing context but also explains that in some instances a host computer 11 (the server in *Thrift*'s client-server architecture) is not needed for at least some speech processing tasks. (Ex. 1015, 3:1-24 (user device "performs all or part of the voice recognition process").) Thus, *Thrift* indicates that it was known before the alleged invention of the '718 patent that tasks could either be allocated to a separate server or performed at the client, depending on particular system needs. (Ex. 1002, ¶¶112-113.)

A POSITA would have understood *Thrift*'s disclosure regarding control unit 10 (the client in *Thrift*'s client-server architecture) performing all or part of a *voice recognition* process to also be applicable to modifying *Cheyer*'s process to have the user's PDA perform all or part of speech recognition *and natural language processing*, because a POSITA would have understood that *Thrift*'s foregoing disclosure is relevant to allocation of tasks in a variety of computational contexts. (*Id.*, ¶114.) In other words, it would have been useful to assign natural language processing to the user's PDA, because natural language processing, like speech recognition, was a task that involved processing data. (*Id.*)

A POSITA would have had reason to consider the teachings of *Thrift* (in the same technical field as *Cheyer*) when implementing *Cheyer*'s process and would have seen that *Thrift* discloses that certain tasks may be assigned to either control unit 10 or to host system 11. (*Id.*, ¶115.) A POSITA would have understood that *Thrift*'s disclosure of control unit 10 performing "all or part" of a voice recognition process (Ex. 1015, 3:1-2, 3:9-10) meant that the choice of which tasks to allocate to the control unit 10 as opposed to host system 11 was determined by system implementation details such as relative resource capabilities. (Ex. 1002, ¶115.) Based on *Thrift*'s disclosure of "control unit 10 perform[ing] *all* voice recognition processes" in one scenario, a POSITA would have recognized the possibility and value of configuring *Cheyer*'s PDA to perform the speech recognition and natural language processing functions disclosed in *Cheyer*. (Ex. 1015, 3:22-23 (emphasis added); Ex. 1002, ¶115.)

For example, a POSITA would have been motivated to make the above modification in order to reduce communications latency, e.g., by eliminating communications to and from a remote server regarding speech recognition and natural language processing. (Ex. 1002, ¶116.) A POSITA would also have been motivated to make this modification to simplify the architecture of *Cheyer*'s system, because with the functions of speech recognition and natural language processing performed at the PDA then a separate speech server would not have been needed for such processing. (*Id.*) A POSITA would have been capable of making this modification, as the choice of a single computer design or a client-server design was a mere choice among a finite number of known alternatives with predictable outcomes. (*Id.*) *KSR*, 550 U.S. at 421.

### 3. Claim 4

#### i) [4.a] "The method of claim 1, further comprising the steps of soliciting additional input from the user, including user interaction in a modality different than the original request;"

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation. (Ex. 1002, ¶117.) At the outset, *Cheyer* discloses several examples in which the user provides additional input beyond just spoken input, including user interaction in a modality different than the original spoken request. (*See, e.g.*, Ex. 1012, 5; Ex. 1002, ¶118.)

*Cheyer* further discloses prompting the user for additional input ("*soliciting* additional input from the user"). (Ex. 1002, ¶119.) For example, *Cheyer* explains where "a user's request is ambiguous or underspecified … the system will … issue a prompt requesting additional information." (*See, e.g.*, *id.*, 6; Ex. 1002, ¶119.)

For example, *Cheyer* discloses prompting the user for an indication (e.g., via a gesture) as to what the user means by "the hotel" in the spoken request. (Ex. 1012, 11, 12.)

#### ii) [4.b] "refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source."

*Cheyer* in combination with *Shwartz* discloses this limitation. (Ex. 1002, ¶120.) For instance, *Cheyer* discloses that in the example of a user input "How far is the restaurant from this hotel" (Ex. 1012, 11), the database request ("navigation query") is refined based upon "a gesture indicating '[this] hotel'" (*id.*, 12), because there is an ambiguity regarding what "this hotel" refers to. (Ex. 1002, ¶120.) *Cheyer* discloses that a "reference resolution agent (RR) … asks for resolution of" a reference such as "[this] hotel" and that "[w]hen the references have been resolved, the domain agent … sends database requests … ." (Ex. 1012, 12.) Thus, *Cheyer* discloses that the database request is refined based upon the additional input from the user that clarifies what the user means by "this hotel" (Ex. 1002, ¶120), and the domain agent sends the refined database request after the ambiguity regarding the reference "this hotel" (Ex. 1012, 11) has been resolved (*id.*, 12). (Ex. 1002, ¶120.) *Cheyer*'s database agent ("the at least one agent") uses the refined database request ("refined navigation query") to retrieve from the remote database location information regarding the hotel specified by the user ("to select a portion of the electronic data source"), so that the distance requested by the user can be calculated. (Ex. 1012, 10 (describing details of database agent); Ex. 1002, ¶120.)

As another example, *Cheyer* discloses that the user may speak "Display the French restaurants within 1 mile of this hotel." (Ex. 1012, 5.) The phrase "this hotel" in this example's spoken query, which is similar to the above-described example involving "the hotel" at page 11 of *Cheyer*, is ambiguous and requires clarification. (Ex. 1002, ¶121.) After the user provides such additional input so that the ambiguity can be resolved, *Cheyer*'s database agent uses a refined database query that takes into account the additional information regarding the identity of the hotel ("the refined navigation query") to select a portion of a database containing maps or "information about available restaurants" (Ex. 1012, 10) relevant to the user's query ("a portion of the electronic data source"). (Ex. 1002, ¶121; *see also* Ex. 1012, 11 ("resolve contextual references such as 'the hotel' … by gestural or direct manipulation commands.").)

To the extent *Cheyer* does not disclose the feature "to select a portion of the electronic data source," it would have been obvious in view of *Shwartz* to implement that feature for at least the same reasons discussed above regarding limitation [1.d]. (*Supra* Section IX.A.1.v; Ex. 1002, ¶122.)

### 4. Claim 6

#### i) [6.a] "The method of claim 1, wherein steps (a)-(d) are performed with respect to multiple users."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses or suggests this limitation. (Ex. 1002, ¶123.) For example, *Cheyer* discloses an application including a user interface that runs on a handheld PDA or a PC (Ex. 1012, Abstract, 4, 6) and further discloses multiple users (*id.*, 1-2 (referring to multiple "users")). A POSITA would have understood that when a plurality of simultaneous users using respective PDAs run *Cheyer*'s application, the method of claim 1, including steps (a)-(d) recited therein, is necessarily performed with respect to multiple users. (Ex. 1002, ¶123.) Indeed, claim 6 does not require that the multiple users share any resources (e.g., remote data sources), and thus amounts to simply having multiple people practice the method of claim 1. At minimum, it would have been obvious to perform steps (a)-(d) with respect to multiple users, e.g., to enable a wider range of people than just one person to be able to use the combined *Cheyer-Thrift-Shwartz* process. (Ex. 1002, ¶123.) A POSITA would have recognized that enabling multiple users to use the combined process would have beneficial, e.g., in order to provide information to more people. (*Id.*)

### 5. Claims 8, 9

#### i) [8.a] "The method of claim 1, wherein the mobile information appliance is a portable computing device."

#### ii) [9.a] "The method of claim 8, wherein the portable computing device is a personal digital assistant."

*Cheyer* combined with *Thrift* and *Shwartz* discloses these limitations. (Ex. 1002, ¶124.) *Cheyer* discloses that the application discussed above for claim 1 runs on a handheld personal digital assistant (PDA), which a POSITA would have understood to be a portable computing device. (Ex. 1012, 4, 6; Ex. 1002, ¶124; *see also* Ex. 1012, Abstract, 12.)

A POSITA would have recognized that the remote control device ("mobile information appliance") in the combined *Cheyer-Thrift-Shwartz* process (discussed above for claim 1) could have additionally been a portable computing device (e.g., PDA), and would have been motivated to implement the device to be both a remote control device and a portable computing device (e.g., PDA). (Ex. 1002, ¶125.) For example, a POSITA would have recognized that the attributes of a remote control device and of a portable computing device (e.g., PDA) were not mutually exclusive, and that these were separate features that could have beneficially have been co-implemented. (*Id.*) Indeed, a POSITA would have been motivated to co-implement both of these features in order to provide a richer feature set for users and to enable a user to perform remote control functionality with an existing device such as his/her portable computing device, e.g., PDA. (*Id.*) Such an implementation would have promoted efficiency, e.g., by using a single device to perform multiple features, and would have been consistent with the knowledge of a POSITA and the expectations of consumers regarding multi-function devices. (*Id.*)

Indeed, it was well-known by the time of the alleged invention that a mobile device could operate as both a PDA and a remote control for a television. (See, e.g., Ex. 1033, 812[8]; *see also* Ex. 1002, ¶126.)

#### 6. Claim 10

##### i) "A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:"

To the extent the preamble of claim 10 is limiting, *Cheyer* discloses the limitations therein for at least the same reasons as presented above regarding the preamble of claim 1. (Ex. 1002, ¶127; *supra* Section IX.A.1.i.)

*Cheyer* discloses an "application" (Ex. 1012, 1-9, 11-12) that "runs on pen-equipped PC's or a Dauphin handheld PDA" (*id.*, 6). *Cheyer* discloses that "[t]o

---

[8] *Konstan* is only cited for claim 1 to demonstrate knowledge of a POSITA and is not relied upon as a reference in this unpatentability ground.

implement the described application, a distributed network of heterogeneous *software* agents was augmented by appropriate functionality for developing synergistic multimodal applications." (*Id.*, 1 (emphasis added).) Therefore, a POSITA would have understood that *Cheyer* discloses a "computer program embodied on a computer readable medium" as claimed. (Ex. 1002, ¶128.)

(*See also infra* Sections IX.A.6.ii-vi regarding the remaining limitations of this claim.)

#### ii) [10.a] "(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;"

*Cheyer* in combination with *Thrift* discloses this limitation for at least the same reasons as presented above regarding limitation [1.a]. (Ex. 1002, ¶129; *supra* Section IX.A.1.ii.)

A POSITA would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.a], and thus *Cheyer* discloses a "code segment" as in limitation [10.a]. (*Supra* Section VIII.B; Ex. 1002, ¶130.) Indeed, even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable and obvious modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("computer processor 12 … controls the overall operation of the system"); Ex. 1002, ¶130.)

#### iii) [10.b] "(b) a code segment that renders an interpretation of the spoken request;"

*Cheyer* discloses this limitation for at least the same reasons as presented above regarding limitation [1.b]. (Ex. 1002, ¶131; *supra* Section IX.A.1.iii.) Moreover, for the same reasons discussed above for limitation [10.a], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [10.b]. (*Supra* Sections VIII.B, IX.A.6.ii; Ex. 1002, ¶132.)

#### iv) [10.c] "(c) a code segment that constructs a navigation query based upon the interpretation;"

*Cheyer* in combination with *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [1.c]. (Ex. 1002, ¶133; *supra* Section IX.A.1.iv.) Moreover, for the same reasons discussed above for limitation [10.a], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [10.c]. (*Supra* Sections VIII.B, IX.A.6.ii; Ex. 1002, ¶134.)

Although *Cheyer* does not expressly describe in detail the limitation "constructs a navigation query," it would have been obvious in view of *Shwartz* to implement that feature in *Cheyer*'s computer program, for at least the same reasons as discussed above for limitation [1.c]. (*Supra* Section IX.A.1.iv; Ex. 1002, ¶135.)

##### v) [10.d] "(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and"

*Cheyer* in combination with *Shwartz*[9] discloses this limitation for at least the same reasons as presented above regarding limitation [1.d]. (Ex. 1002, ¶136; *supra* Section IX.A.1.v.) Moreover, for the same reasons discussed above for limitation [10.a], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [10.d]. (*Supra* Sections VIII.B, IX.A.6.ii; Ex. 1002, ¶137.)

---

[9] As discussed above for limitation [10.c], it would have been obvious in view of *Shwartz* to modify *Cheyer*'s computer program to construct a "navigation query." It would also have been obvious to configure the combined *Cheyer-Shwartz* computer program to implement the "navigation query" feature in limitation [10.d] and claim 13. (Ex. 1002, ¶136 n.5.)

To the extent *Cheyer* does not disclose "select a portion of the electronic data source," it would have been obvious in view of the combined teachings of *Cheyer* and *Shwartz* to implement this feature in *Cheyer*'s computer program, for at least the same reasons as discussed above for limitation [1.d]. (*Supra* Section IX.A.1.v; Ex. 1002, ¶138.)

#### vi) [10.e] "(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user."

*Cheyer* discloses this limitation for at least the same reasons as presented above regarding limitation [1.e]. (Ex. 1002, ¶139; *supra* Section IX.A.1.vi.) Moreover, for the same reasons discussed above for limitation [10.a], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [10.e]. (*Supra* Sections VIII.B, IX.A.6.ii; Ex. 1002, ¶140.)

### 7. Claim 12

#### i) [12.a] "The computer program of claim 10, wherein the rendering of the interpretation of the spoken request is performed by the mobile information appliance."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claims 2 and 3. (Ex. 1002, ¶141; *supra* Section IX.A.2.)

### 8. Claim 13

#### i) [13.a] "The computer program of claim 10, further comprising a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;"

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.a]. (Ex. 1002, ¶142; *supra* Section IX.A.3.i.) Moreover, for the same reasons discussed above for limitation [10.a], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [13.a]. (*Supra* Sections VIII.B, IX.A.6.ii; Ex. 1002, ¶143.)

#### ii) [13.b] "a code segment that refines the navigation query, based upon the additional input; and a code segment that uses the refined navigation query to select a portion of the electronic data source."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.b]. (Ex. 1002, ¶144; *supra* Section IX.A.3.ii.) Moreover, for the same reasons discussed above for limitation [10.a], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [13.b]. (*Supra* Sections VIII.B, IX.A.6.ii; Ex. 1002, ¶145.)

### 9. Claim 15

#### i) [15.a] "The computer program of claim 10, wherein code segments (a)-(d) are executed with respect to multiple users."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claim 6. (Ex. 1002, ¶146; *supra* Section IX.A.4.)

### 10. Claims 17, 18

#### i) [17.a] "The computer program of claim 10, wherein the mobile information appliance is a portable computing device."

#### ii) [18.a] "The computer program of claim 17, wherein the portable computing device is a personal digital assistant."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses these limitations for at least the same reasons as presented above regarding claims 8 and 9. (Ex. 1002, ¶147; *supra* Sections IX.A.5.i-ii.)

### 11. Claim 19

#### i) "A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:"

To the extent the preamble of claim 19 is limiting, *Cheyer* discloses the limitations therein for at least the same reasons as presented above regarding the preamble of claim 1. (Ex. 1002, ¶148; *supra* Section IX.A.1.i.)

In addition to disclosing a "method" as recited in claim 1 (*supra* Section IX.A.1.i), *Cheyer* also expressly discloses a "system" utilizing an application that runs on a PC or PDA (Ex. 1012, 6), and thus discloses a "system" as recited in the preamble of claim 19. (Ex. 1002, ¶149; *see also* Ex. 1012, 4, 6, 12.)

(*See also infra* Sections IX.A.11.ii-vi regarding the remaining limitations of this claim.)

#### ii) [19.a] "(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;"[10]

*Cheyer* in combination with *Thrift* discloses this limitation for at least the same reasons as presented above regarding the preamble of claim 1 and limitation [1.a]. (Ex. 1002, ¶150; *supra* Sections IX.A.1.i-ii.)

#### iii) [19.b] "(b) spoken language processing logic, operable to render an interpretation of the spoken request;"

*Cheyer* discloses this limitation for at least the same reasons as presented above regarding limitation [1.b]. (Ex. 1002, ¶151; *supra* Section IX.A.1.iii.)

---

[10] Limitation [19.a] defines sufficient structure ("a portable remote control device or a set-top box for a television") to avoid invoking 35 U.S.C. § 112 ¶ 6.

Additionally, because *Cheyer* discloses an application implemented in software (*supra* Section IX.A.6.i) and a POSITA would have understood that such software runs on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [19.b], and for the reasons discussed above for limitation [1.b], *Cheyer* discloses a "spoken language processing logic, operable to" perform the functionality recited in limitation [19.b]. (*Supra* Sections VIII.B, IX.A.1.iii, IX.A.6.iii; Ex. 1002, ¶152.) Indeed, even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a processor to implement logic was routine and commonplace at the time of the alleged invention, and would have been a predictable and obvious modification. (Ex. 1013, 4:11-62, 6:29-30; Ex. 1002, ¶152.)

#### iv) [19.c] "(c) query construction logic, operable to construct a navigation query based upon the interpretation;"

*Cheyer* in combination with *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [1.c]. (Ex. 1002, ¶153; *supra* Section IX.A.1.iv.) Moreover, for the same reasons discussed above for limitation [19.b], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [19.c]. (*Supra* Sections VIII.B, IX.A.11.iii; Ex. 1002, ¶154.)

Although *Cheyer* does not expressly describe in detail the limitation "construct a navigation query," it would have been obvious in view of *Shwartz* to implement that feature in *Cheyer*'s system, for at least the same reasons as discussed above for limitation [1.c]. (*Supra* Section IX.A.1.iv; Ex. 1002, ¶155.)

#### v) [19.d] "(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and"

*Cheyer* in combination with *Shwartz*[11] discloses this limitation for at least the same reasons as presented above regarding limitation [1.d]. (Ex. 1002, ¶156; *supra* Section IX.A.1.v.) Moreover, for the same reasons discussed above for limitation [19.b], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [19.d]. (*Supra* Sections VIII.B, IX.A.11.iii; Ex. 1002, ¶157.)

---

[11] As discussed above for limitation [19.c], it would have been obvious in view of *Shwartz* to modify *Cheyer*'s system to construct a "navigation query." It would also have been obvious to configure the combined *Cheyer*-*Shwartz* system to implement the "navigation query" feature in limitation [19.d] and claim 22. (Ex. 1002, ¶156 n.6.)

#### vi) [19.e] "(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user."

*Cheyer* alone and/or in combination with *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [1.e]. (Ex. 1002, ¶158; *supra* Section IX.A.1.vi.) A POSITA would have understood that *Cheyer* alone and/or in combination with *Shwartz* necessarily discloses an *electronic communications infrastructure* for performing the transmitting of limitation [19.e]. (Ex. 1002, ¶158.) A POSITA would have had this understanding because without an electronic communications infrastructure, a system like that disclosed in *Cheyer* and *Shwartz*, which involve retrieving information from a remote system (*supra* Sections IX.A.1.i, vi, IX.A.11.i) would not have been possible. (Ex. 1002, ¶158.) Indeed, an electronic communications infrastructure was a necessary component of a remote server (e.g., web server) such as disclosed by *Cheyer* in the context of a Web-based data source. (*Supra* Section IX.A.1.i; Ex. 1012, Abstract ("access to existing data sources including the World Wide Web"), 6 ("a mobile system that provides a synergistic pen/voice interface to remote databases"); Ex. 1002, ¶158.)

To the extent the claimed "electronic communications infrastructure for transmitting …" is construed to require software running on a microprocessor configured to perform transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user (*see supra* n.3), *Cheyer* discloses this limitation for at least the same reasons presented above regarding limitation [10.e]. (*Supra* Section IX.A.6.vi; Ex. 1002, ¶159.)

### 12. Claim 21

#### i) [21.a] "The system of claim 19, wherein the spoken language processing logic renders the interpretation of the spoken request at the mobile information appliance."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claims 2 and 3. (Ex. 1002, ¶160; *supra* Section IX.A.2.)

### 13. Claim 22

#### i) [22.a] "The system of claim 19, further comprising user interaction logic operable to solicit additional input from the user, including user interaction in a modality different than the original request; and"

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.a]. (Ex. 1002, ¶161; *supra* Section IX.A.3.i.) Moreover, for the same reasons discussed above for limitation [19.b], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [22.a]. (*Supra* Sections VIII.B, IX.A.11.iii; Ex. 1002, ¶162.)

#### ii) [22.b] "query refining logic operable to refine the navigation query based upon the additional input; wherein the navigation logic users[12] the refined navigation query to select a portion of the electronic data source."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.b]. (Ex. 1002, ¶¶163, 165; *supra* Section IX.A.3.ii.) Moreover, for the same reasons discussed above for limitation [19.b], *Cheyer* discloses, or renders obvious in view of *Shwartz*, the use of a processor to implement the functionality recited in limitation [22.b]. (*Supra* Sections VIII.B, IX.A.11.iii; Ex. 1002, ¶164.)

### 14. Claim 24

#### i) [24.a] "The system of claim 19, wherein the system operates with respect to multiple users."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claim 6. (Ex. 1002, ¶166; *supra* Section IX.A.4.)

---

[12] For the purposes of this proceeding, Petitioner assumes that claim 22 contains a typographical error and was intended to recite "uses" instead of "users." Petitioner reserves the right to assert invalidity of claim 22 under 35 U.S.C. § 112 in other proceedings.

#### 15. Claims 26, 27

##### i) [26.a] "The system of claim 19, wherein the mobile information appliance is a portable computing device."

##### ii) [27.a] "The system of claim 26, wherein the portable computing device is a personal digital assistant."

*Cheyer* in combination with *Thrift* and *Shwartz* discloses these limitations for at least the same reasons as presented above regarding claims 8 and 9. (Ex. 1002, ¶167; *supra* Sections IX.A.5.i-ii.)

### B. Ground 2: *Cheyer*, *Shwartz*, *Thrift*, and *Dureau* Render Obvious Claims 2, 11, and 20

#### 1. Claim 2

##### i) [2.a] "The method of claim 1, wherein the step of rendering the interpretation of the spoken request is performed [at the one or more network servers]."[13]

*Cheyer* combined with *Thrift*, *Shwartz*, and *Dureau*, discloses or suggests this limitation. (Ex. 1002, ¶169.)

*Cheyer* discloses a "server machine which will manage … natural language processing and speech recognition for the application." (Ex. 1012, 6; *see also id.*, 4, 11.) While *Cheyer* does not expressly disclose that the server at which the data source is located according to the preamble of claim 1 also performs speech recognition and natural language processing ("the step of rendering the

---

[13] *Supra* n.7.

interpretation of the spoken request”), to the extent claim 2 is interpreted to require such an arrangement, it would have been obvious in view of *Dureau* to configure the combined *Cheyer-Shwartz-Thrift* process to implement such features. (Ex. 1002, ¶170.)

*Dureau* “relates generally to interactive television systems” (Ex. 1016, 1:8-12) and discloses voice input to a set-top box coupled to a television. (*Id.*, Abstract, 10:56-11:1, FIG. 1 (reproduced below); Ex. 1002, ¶171.)

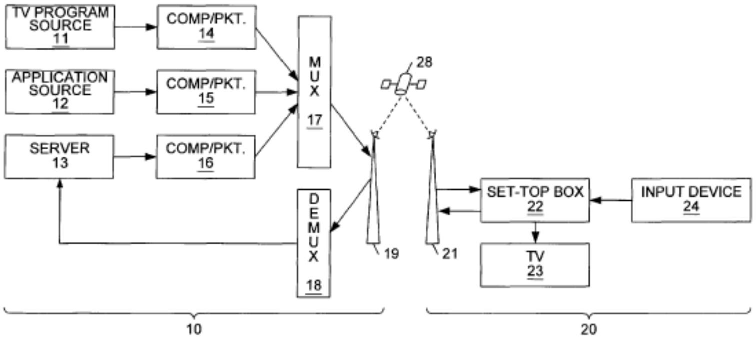


FIG. 1

(Ex. 1016, FIG. 1 (showing set-top box 22 connected to television 23.)

Because *Dureau*, like *Cheyer*, discloses that a user provides voice input that is processed by voice recognition software, a POSITA would have had reason to consider the teachings of *Dureau* when implementing the *Cheyer-Shwartz-Thrift* process. (Ex. 1002, ¶172.) Having looked to *Dureau*, a POSITA would have seen that *Dureau* discloses transmitting a user’s speech input to a server, where it is interpreted, and further discloses performing applications relating to the speech input at the server. (Ex. 1016, Abstract, 2:49-62, 3:39-44, 9:59-10:3, 10:46-55, 10:65-11:3; Ex. 1002, ¶172.)

Based on *Dureau*'s disclosures regarding a server that is equipped with a voice recognition application and that performs applications using speech input, a POSITA would have been motivated to modify the combined *Cheyer-Thrift-Shwartz* process so that the server at which the data source is located as in the preamble of claim 1 also performs speech recognition and natural language processing. (Ex. 1002, ¶173.) A POSITA would have known based on *Dureau* that such a configuration was possible, and he/she would have been motivated to implement the data source at the same server that performs speech recognition and natural language processing in order to achieve an efficient implementation. (*Id.*) Such an implementation would have been a mere combination of known components and technologies, according to known methods, to achieve predictable results. (*Id.*) *KSR*, 550 U.S. at 416.

### 2. Claims 11, 20

#### i) [11.a] "The computer program of claim 10, wherein the rendering of the interpretation of the spoken request is performed at the one or more network servers."

#### ii) [20.a] "The system of claim 19, wherein the spoken language processing logic renders the interpretation of the spoken request at the one or more network servers."

*Cheyer* in combination with *Thrift*, *Shwartz*, and *Dureau* discloses or suggests these limitations for at least the same reasons as presented above regarding claim 2 in Ground 2. (Ex. 1002, ¶¶174-175; *supra* Section IX.B.1.)

## C. Ground 3: *Cheyer*, *Shwartz*, *Thrift*, and *Johnson* Render Obvious Claims 4, 13, and 22

Patent Owner may argue that claims 4, 13, and 22 should be interpreted to require that the soliciting of additional input and refining of the navigation query recited in those claims must be for a different navigation query than the one recited in claim limitations [1.d], [10.d], and [19.d]. To the extent such an interpretation is applied, claims 4, 13, and 22 nonetheless would have been obvious in view of *Johnson*'s additional disclosure. (Ex. 1002, ¶176.) Indeed, such an approach was well within the skill of a POSITA and a mere design choice. (*Id.*)

### 1. Claim 4

#### i) Claim limitation [4.a]

To the extent that *Cheyer*, *Thrift*, and *Shwartz* may not explicitly teach limitation [4.a], it would have been obvious in view of *Johnson* to modify the combined *Cheyer*-*Thrift*-*Shwartz* process (discussed above for claim 1) to include such features. (Ex. 1002, ¶177.) *Johnson*, which is directed to "a multimodal natural language interface [that] interprets user requests," is in the same technical field as *Cheyer*. (Ex. 1002, ¶178; Ex. 1014, Abstract.) A POSITA would have had reason to consider the teachings of *Johnson* for enhancing or augmenting the capabilities of the combined *Cheyer*-*Thrift*-*Shwartz* method, because *Cheyer*, *Thrift*, *Shwartz*, and *Johnson* are all directed to servicing user requests that are provided via an interface that includes natural language input. (Ex. 1002, ¶178.)

*Johnson* discloses that in the example of a database query for "Joe Smith's telephone number," there could be "two Joe Smiths in the database," so that "there is an ambiguity that must be clarified before a final response can be generated." (Ex. 1014, 5:7-18; *see also id.*, Abstract, 4:9-12, FIG. 4 (reproduced below).)

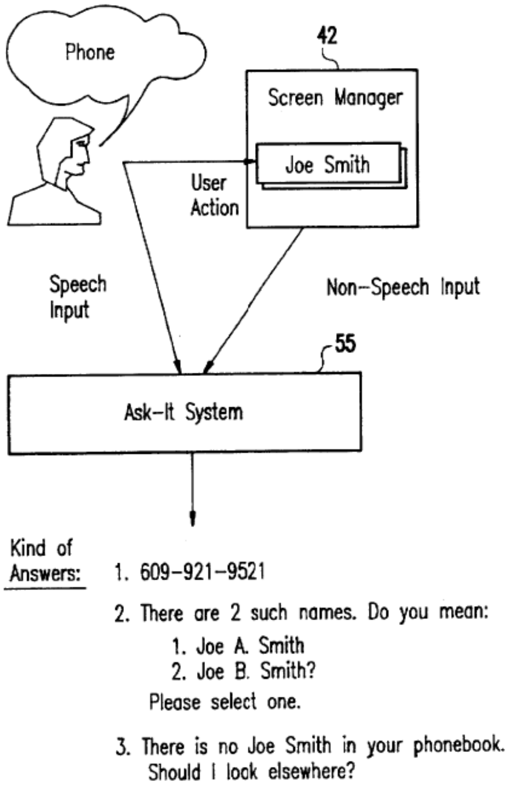


FIG.4

Thus, as shown in Figure 4, if there is an ambiguity, *Johnson*'s system asks the user to "select one" of the possibilities or indicate whether to look elsewhere. (Ex. 1014, FIG. 4; Ex. 1002, ¶¶179-180.)

In view of *Johnson*'s disclosure of seeking clarification regarding an ambiguous situation in which two possible results are present, a POSITA would have been motivated to modify the combined *Cheyer*-*Thrift*-*Shwartz* process to clarify any ambiguity in a similar manner, and thus would have been motivated to solicit additional input from the user regarding such clarification, to provide the user with desired information in a user-friendly and convenient manner. (Ex. 1002, ¶181.)

In view of *Johnson*'s disclosure of "provid[ing] a choice to the user … in a pop-up window, and request[ing] the user to select one of the choices" (Ex. 1014, 5:11-12), a POSITA would have been motivated to configure the combined process to include user interaction in a modality such as via a selection from a pop-up window in a graphical user interface without using voice input ("a modality different than the original request"), because such a skilled person would have recognized that providing the user with an ability to select a choice from a pop-up window by, for example, touching or clicking the choice would have been a convenient, simple, and user-friendly implementation that would have enabled a wider range of input options for the user. (Ex. 1002, ¶182.)

Indeed, *Cheyer* and *Johnson* encourage such multimodal interaction, disclosing several examples in which the user provides non-spoken input. (*See, e.g.*, Ex. 1012, 1, 4, 5; Ex. 1014, Abstract, 2:21-22, 3:37-42, 3:44-46, 3:49-51; Ex. 1002, ¶183.) Furthermore, input modalities other than speech input were well known long before the alleged invention of the '718 patent. (Ex. 1002, ¶183.)

In view of *Cheyer*'s and *Johnson*'s encouragement of multimodal input, a POSITA would have been motivated to modify the combined *Cheyer-Thrift-Shwartz* process to implement the features of limitation [4.a]. (*Id.*, ¶184.) This modification would have been a mere combination of known components and technologies, according to known methods, to obtain predictable results. (*Id.*) *KSR*, 550 U.S. at 416.

#### ii) Claim limitation [4.b]

To the extent that *Cheyer*, *Thrift*, and *Shwartz* may not explicitly teach refining the navigation query, based upon the additional input, and using the refined navigation query to select a portion of the electronic data source, it would have been obvious in view of *Johnson* to modify the combined *Cheyer*-*Thrift*-*Shwartz* process to implement such features. (Ex. 1002, ¶185.)

*Johnson* explicitly recognizes that ambiguities may be detected after an electronic data source is accessed, necessitating refinement of a navigation query and searching of an electronic data source after a first navigation query already searches the data source. (Ex. 1002, ¶185.) For example, as discussed above in Section IX.C.1.i, *Johnson* discloses that in the example of a database query for "Joe Smith's telephone number," there could be "two Joe Smiths [found] in the database" after searching the database, so that "there is an ambiguity that must be clarified before a final response can be generated." (*Supra* Section IX.C.1.i; Ex. 1014, 5:7-18.) Thus, *Johnson* discloses requesting the user to select one of a plurality of choices or to specify whether a search should be conducted elsewhere. (Ex. 1014, FIG. 4; Ex. 1002, ¶185.) As a result of the user's selection, *Johnson*'s system can find and present to the user the phone number that the user requested ("a portion of the electronic data source"). (Ex. 1002, ¶185.)

A POSITA would have been motivated to include in the combined *Cheyer-Thrift-Shwartz* process features such as those disclosed in *Johnson* regarding refining the navigation query after a database is initially searched based upon additional input and using the refined navigation query to select a portion of a database ("the electronic data source"), in order to enable the combined method and system to be able to handle situations where a user's request results in multiple, ambiguous hits or no hits at all. (*Id.*, ¶186.) This would have been a simple modification for a POSITA to make, as it would have been merely a combination of known elements, according to known methods, to yield predictable results. (*Id.*; *see also supra* Section IX.A.1.v.) *KSR*, 550 U.S. at 416. Indeed, a POSITA would have recognized that accessing and selecting a portion of an electronic data source with a refined navigation query would have involved substantially the same operations as compared to accessing and selecting a portion of an electronic data source with an original navigation query. (Ex. 1002, ¶186.)

Indeed, a POSITA would have recognized the existence of two options for leveraging the user's clarification in *Johnson* to obtain the phone number of the Joe Smith intended by the user: (a) access the database with a search query specifying "Joe Smith" and obtain an indication that there are two Joe Smiths in the database, without obtaining at that time the phone number for each Joe Smith (such that the phone number for the user-intended Joe Smith must later be retrieved from the database after the user's clarification); and (b) access the database with a search query specifying "Joe Smith" and obtain an indication that there are two Joe Smiths in the database, along with their respective phone numbers (such that upon the user's clarification, the user-intended phone number can simply be used without further accessing the database). (*Id.*, ¶187.)

A POSITA would have recognized that configuring the combined *Cheyer-Thrift-Shwartz-Johnson* process to use the refined navigation query to select a portion of the electronic data source would have constituted a mere design choice among a finite number of known alternatives (e.g., the foregoing two options, which are not mutually exclusive, as a POSITA would have recognized that ambiguities could be resolved both before and after accessing the database), each having predictable outcomes (e.g., ultimately obtaining from the database the phone number of the user-intended Joe Smith). (*Id.*, ¶188.) *KSR*, 550 U.S. at 421.

Moreover, to the extent *Johnson* does not disclose the feature "to select a portion of the electronic data source," it would have been obvious in view of *Shwartz* to implement that feature in the combined process. (*Supra* Section IX.A.1.v; Ex. 1002, ¶189.)

#### 2. Claim 13

##### i) Limitations [13.a], [13.b]

*Cheyer* in combination with *Thrift*, *Shwartz*, and *Johnson* discloses or suggests these limitations for at least the same reasons as presented above. (*Supra* Sections IX.A.6, IX.A.8.i-ii (citations and analysis regarding "code segment[s]"), IX.C.1; Ex. 1002, ¶190.)

#### 3. Claim 22

##### i) Limitations [22.a], [22.b]

*Cheyer* in combination with *Thrift*, *Shwartz* and *Johnson* discloses or suggests these limitations for at least the same reasons as presented above. (*Supra* Sections IX.A.11, IX.A.13.i-ii (citations and analysis regarding "logic operable to"), IX.C.1; Ex. 1002, ¶191.)

### D. Ground 4: *Cheyer*, *Shwartz*, *Thrift*, and *Simmers* Render Obvious Claims 5, 7, 14, 16, 23, and 25

#### 1. Claims 5, 7

##### i) [5.a] "The method of claim 1, wherein the data link includes a cellular telephone system."

##### ii) [7.a] "The method of claim 1, wherein the mobile information appliance is a wireless telephone."

As discussed above for claim 1, *Cheyer* discloses a data link between the user's mobile device and a remote server. (*Supra* Section IX.A.1.i; Ex. 1002, ¶193.) While *Cheyer*, *Thrift*, and *Shwartz* do not expressly disclose a data link including a cellular telephone system, it would have been obvious in view of *Simmers* to configure the combined *Cheyer-Thrift-Shwartz* process to implement this feature. (Ex. 1002, ¶193.)

*Cheyer* discloses that the mobile device can be a PDA (Ex. 1012, 4, 6), and a POSITA implementing *Cheyer*'s process would have recognized the desirability of incorporating cellular telephone functionality into a PDA. (Ex. 1002, ¶194.) For example, *Simmers* discloses "dual-function information devices such as a cellular phone with PDA." (Ex. 1017, 1:47-48; *see also id.*, 1:12-15.)

A POSITA would have recognized that the mobile information appliance (e.g., a PDA with remote control functionality) in the combined *Cheyer-Thrift-Shwartz* process could have also included wireless telephone functionality, and would have been motivated to implement both in the device. (Ex. 1002, ¶195.) Moreover, it was well known before the alleged invention of the '718 patent that a cellular phone (e.g., as disclosed by *Simmers*) was used for communicating across a *cellular telephone system*. (Ex. 1002, ¶196.)

A POSITA would have recognized the value of implementing a cellular telephone system (which *Simmers*'s cellular-enabled PDA would have used) to achieve a data link between *Cheyer*'s mobile device and remote data source. (Ex. 1002, ¶197.) For example, a POSITA would have recognized that a cellular telephone system was a known system for communicating between a mobile device and a remote computer on the Web, and that *Cheyer* similarly discloses communications between a mobile device and a data source on the web (*supra* Section IX.A.1.i). (Ex. 1002, ¶197.) In view of *Simmers*'s teachings, a POSITA would have been motivated and been capable of modifying the combined *Cheyer-Thrift-Shwartz* process so that *Cheyer*'s data link discussed above for the preamble of claim 1 includes a cellular telephone system, as recited in claim 5. (*Id.*)

This would have been a mere combination of known components and technologies (e.g., *Cheyer*'s disclosure of communication between a user's mobile device and a remote data source to provide the user with desired data from the data source, and *Simmers*'s disclosure of a cellular telephone system), according to known methods (e.g., a POSITA would have known how to implement a cellular telephone system to achieve *Cheyer*'s communication between a mobile device and a remote data source), to obtain predictable results (e.g., communication between two devices using a known networking technology). (*Id.*, ¶¶198-199.) *KSR*, 550 U.S. at 416.

A POSITA would further have been motivated in view of the foregoing references to configure *Cheyer*'s mobile device of the user ("mobile information appliance") to be a wireless telephone, as recited in claim 7. (Ex. 1002, ¶200.) For example, a POSITA would have known that a wireless telephone was typically used with a cellular telephone system to provide portability and mobile access to data sources. (*Id.*)

Such a configuration would have been a mere combination of known components and technologies (e.g., a known cellular telephone system and a wireless telephone that was known to be used with such a cellular telephone system, *Cheyer*'s disclosure of a mobile device such as a PDA, and *Simmers*'s disclosure of a cellular-enabled PDA), according to known methods (e.g., a POSITA knew how to configure a device to be a wireless telephone), to achieve predictable results (e.g., providing a user with a wireless telephone). (Ex. 1012, 4, 6; Ex. 1017, 1:47-48; Ex. 1002, ¶201.) *KSR*, 550 U.S. at 416.

Indeed, a PDA's flexibility and expandability were well-known by the time of the alleged invention, and it was well-known that a PDA could operate as both a cellular phone (Ex. 1017, 1:47-48) and a remote control (Ex. 1033, 812). (*See also* Ex. 1002, ¶202.)

### 2. Claims 14, 16, 23, 25

#### i) [14.a] "The computer program of claim 10, wherein the data link includes a wireless telephone system."

#### ii) [16.a] "The computer program of claim 10, wherein the mobile information appliance is a wireless telephone."

#### iii) [23.a] "The system of claim 19, wherein the data link includes a cellular telephone system."

#### iv) [25.a] "The system of claim 19, wherein the mobile information appliance is a wireless telephone."

*Cheyer* in combination with *Thrift*, *Shwartz*, and *Simmers* discloses these limitations for at least the same reasons as presented above regarding claims 5 and 7. (Ex. 1002, ¶203; *supra* Sections IX.D.1.i-ii.) It would have been obvious to implement a *wireless* telephone system as recited in claim 14 for similar reasons as discussed above for claim 5 regarding implementing a *cellular* telephone system, because a cellular telephone system was a type of wireless telephone system. (Ex. 1002, ¶203.)

## X. IPR SHOULD BE INSTITUTED ON ALL GROUNDS

In Ground 1, Petitioner relies on *Cheyer*, *Shwartz*, and *Thrift* to address claims 4, 13, and 22. In Ground 3, Petitioner addresses those claims based on the additional disclosures in *Johnson*. While *Cheyer* and *Johnson* both disclose soliciting additional input beyond a spoken request, *Cheyer* discloses refining a database query before it is used to retrieve information from a database, whereas

*Johnson* discloses refining a query after already accessing the database. Depending on Patent Owner's positions and/or the Board's interpretation of the references and/or the claims, either Ground 1 or Ground 3 may have strengths or weaknesses relative to the other. Both these grounds, as well as Grounds 2 and 4 (which introduce secondary references for certain dependent claims), should be instituted in order to enable fuller development of the record.

## XI. CONCLUSION

For the reasons given above, Petitioner requests institution of IPR for claims 1-27 of the '718 patent based on each of the grounds specified in this petition.

Respectfully submitted,

Dated: January 12, 2018 By: /Naveen Modi/
Naveen Modi (Reg. No. 46,224)

# CERTIFICATE OF COMPLIANCE

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that the foregoing Petition for *Inter Partes* Review of U.S. Patent No. 6,757,718 contains, as measured by the word-processing system used to prepare this paper, 13,975 words. This word count does not include the items excluded by 37 C.F.R. § 42.24 as not counting towards the word limit.

Respectfully submitted,

Dated: January 12, 2018 By: /Naveen Modi/
Naveen Modi (Reg. No. 46,224)
Counsel for Petitioner

# CERTIFICATE OF SERVICE

I hereby certify that on January 12, 2018, I caused a true and correct copy of the foregoing Petition for *Inter Partes* Review of U.S. Patent No. 6,757,718 and supporting exhibits to be served via express mail on the Patent Owner at the following correspondence address of record as listed on PAIR:

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

By: /Naveen Modi/
Naveen Modi (Reg. No. 46,224)

US006757718B1

(12) **United States Patent**
**Halverson et al.**

(10) **Patent No.: US 6,757,718 B1**
(45) **Date of Patent: Jun. 29, 2004**

(54) **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT**

(75) Inventors: **Christine Halverson**, San Jose, CA (US); **Luc Julia**, Menlo Park, CA (US); **Dimitris Voutsas**, Thessaloniki (GR); **Adam Cheyer**, Palo Alto, CA (US)

(73) Assignee: **SRI International**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) bydays.days.

(21) Appl. No.: **09/608,872**

(22) Filed: **Jun. 30, 2000**

**Related U.S. Application Data**

(63) Continuation of application No. 09/524,095, filed on Mar. 13, 2000, which is a continuation-in-part of application No. 09/225,198, filed on Jan. 5, 1999.

(60) Provisional application No. 60/124,720, filed on Mar. 17, 1999, provisional application No. 60/124,719, filed on Mar. 17, 1999, and provisional application No. 60/124,718, filed on Mar. 17, 1999.

(51) **Int. Cl.**$^7$ .................................................. **G06F 15/16**

(52) **U.S. Cl.** ........................ **709/218**; 709/202; 709/217; 709/219; 709/227; 704/257

(58) **Field of Search** ................................ 709/202, 218, 709/217, 219, 227; 707/5, 3, 4; 704/257, 270.1, 275, 246

(56) **References Cited**

U.S. PATENT DOCUMENTS

| | | | | |
|---|---|---|---|---|
| 5,197,005 | A | 3/1993 | Shwartz et al. | 364/419 |
| 5,386,556 | A | 1/1995 | Hedin et al. | 395/600 |
| 5,434,777 | A | 7/1995 | Luciw | 364/419.13 |
| 5,519,608 | A | 5/1996 | Kupiec | 364/419.08 |
| 5,608,624 | A | 3/1997 | Luciw | 395/794 |
| 5,721,938 | A | 2/1998 | Stuckey | 395/754 |
| 5,729,659 | A | 3/1998 | Potter | 395/2.79 |
| 5,748,974 | A | 5/1998 | Johnson | 395/759 |
| 5,774,859 | A | 6/1998 | Houser et al. | 704/275 |
| 5,794,050 | A | 8/1998 | Dahlgren et al. | 395/708 |
| 5,802,526 | A | 9/1998 | Fawcett et al. | 707/104 |
| 5,805,775 | A | 9/1998 | Eberman et al. | 395/12 |
| 5,855,002 | A | 12/1998 | Armstrong | 704/270 |
| 5,890,123 | A | 3/1999 | Brown et al. | 704/275 |
| 5,963,940 | A | 10/1999 | Liddy et al. | 707/5 |

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| EP | 0 867 861 | 9/1998 | G10L/5/06 |
| WO | 99/50826 | 10/1999 | G10L/3/00 |
| WO | 00/05638 | 2/2000 | |

OTHER PUBLICATIONS

International Search Report, Intl Appl No. PCT/US01/07987.

Stent, Amanda et al., "The CommandTalk Spoken Dialogue System", SRI International.

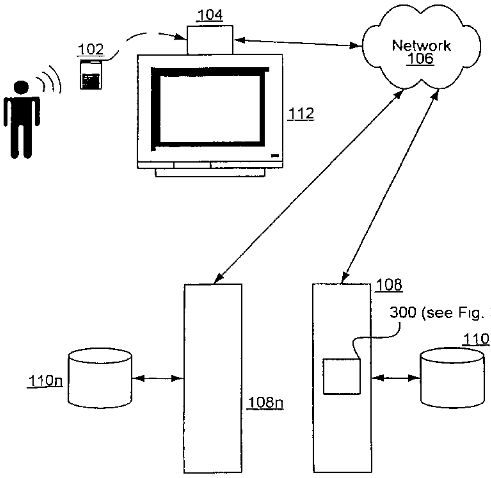
(List continued on next page.)

*Primary Examiner*—Frantz B Jean
(74) *Attorney, Agent, or Firm*—Moser, Patterson & Sheridan, LLP; Kin-Wah Tong

(57) **ABSTRACT**

A system, method, and article of manufacture are provided for navigating an electronic data source by means of spoken language where a portion of the data link between a mobile information appliance of the user and the data source utilizes wireless communication. When a spoken input request is received from a user who is using the mobile information appliance, it is interpreted. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is transmitted to the mobile information appliance.

**27 Claims, 7 Drawing Sheets**



U.S. PATENT DOCUMENTS

| | | | | | |
|---|---|---|---|---|---|
| 6,003,072 | A | | 12/1999 | Gerritsen et al. | 709/218 |
| 6,016,476 | A | | 1/2000 | Maes et al. | 705/1 |
| 6,026,388 | A | | 2/2000 | Liddy et al. | 707/1 |
| 6,102,030 | A | | 8/2000 | Brown et al. | 704/275 |
| 6,173,279 | B1 | * | 1/2001 | Levin et al. | 707/5 |
| 6,192,338 | B1 | * | 2/2001 | Haszto et al. | 704/257 |
| 6,314,365 | B1 | * | 11/2001 | Smith | 340/988 |
| 6,317,684 | B1 | * | 11/2001 | Roeseler et al. | 340/990 |
| 6,349,257 | B1 | * | 2/2002 | Liu et al. | 340/5.6 |
| 6,353,661 | B1 | * | 3/2002 | Bailey, III | 379/88.17 |

OTHER PUBLICATIONS

Moore, Robert et al., "CommandTalk: A Spoken–Language Interface for Battlefield Simulations", Oct. 23, 1997, SRI International.

Dowding, John et al., "Interpreting Language in Context in CommandTalk", Feb. 5, 1999, SRI International.

http://www.ai.sri.com/~oaa/infowiz.html, InfoWiz: An Animated Voice Interactive Information System, May 8, 2000.

Dowding, John, "Interleaving Syntax and Semantics in an Efficient Bottom–up Parser", SRI International.

Moore, Robert et al., "Combining Linguistic and Statistical Knowledge Sources in Natural–Language Processing for ATIS", SRI International.

Dowding, John et al., "Gemini: A Natural Language System For Spoken–Language Understanding", SRI International.

* cited by examiner

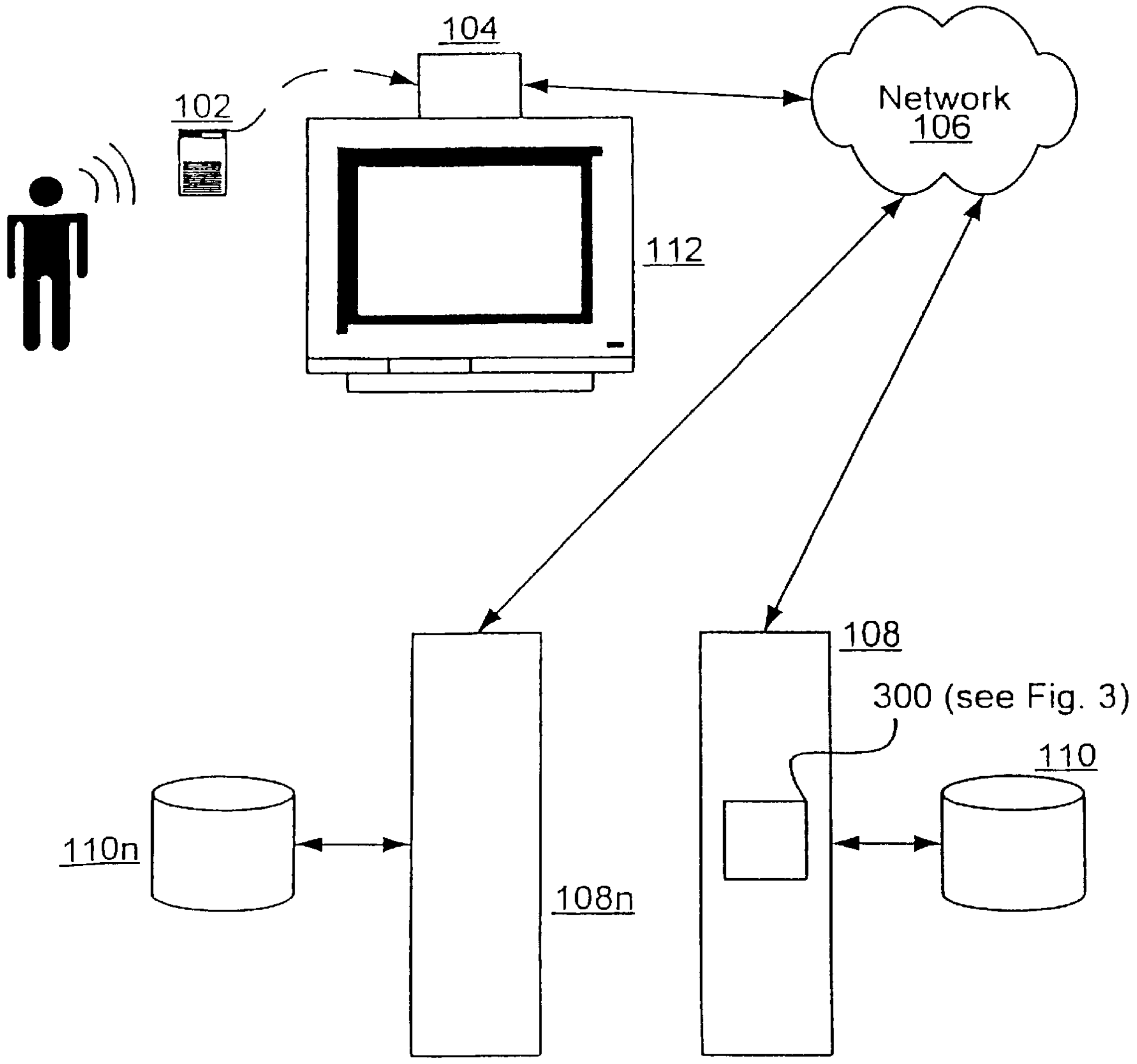


Fig. 1a

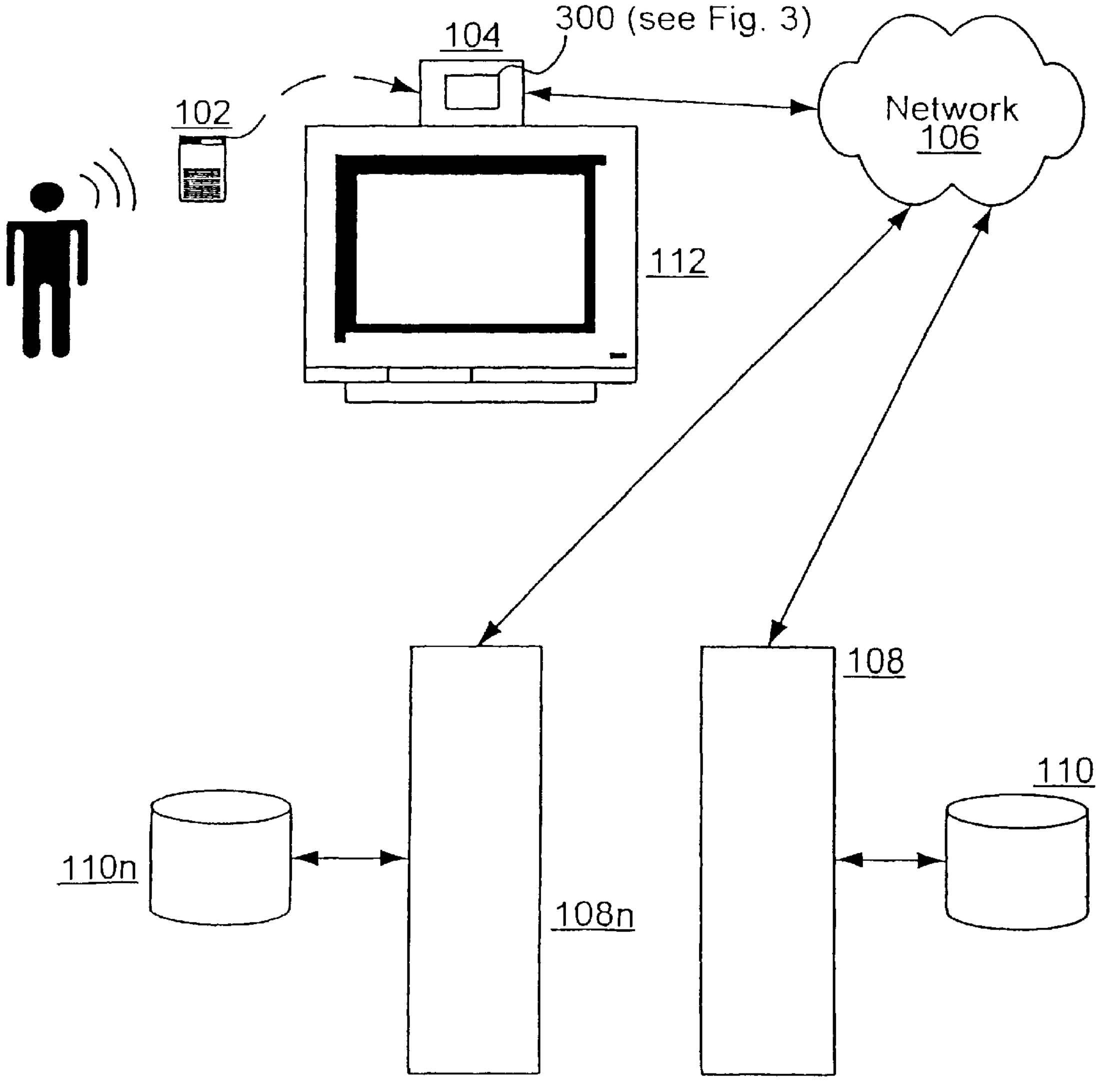


Fig. 1b

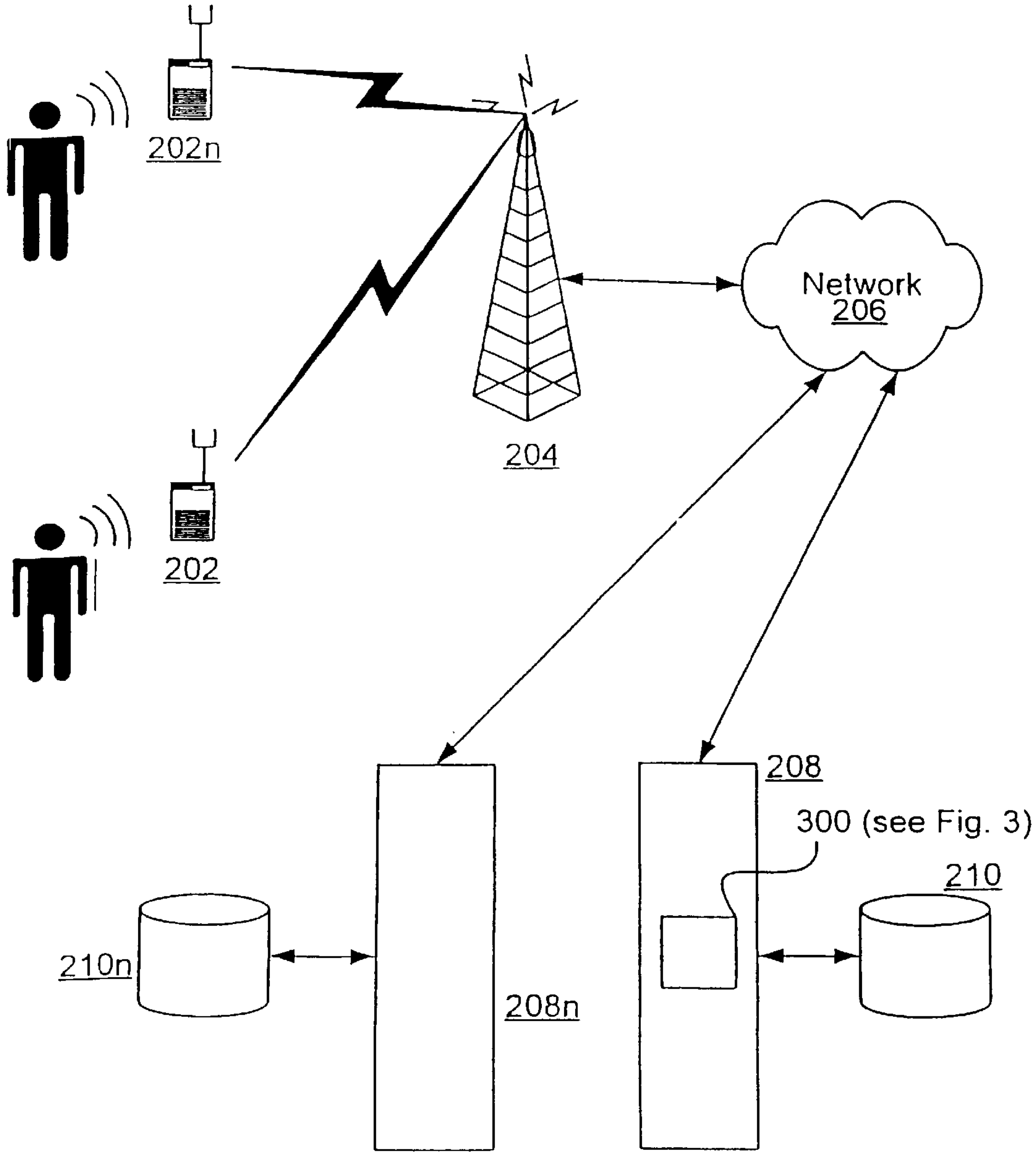


Fig. 2

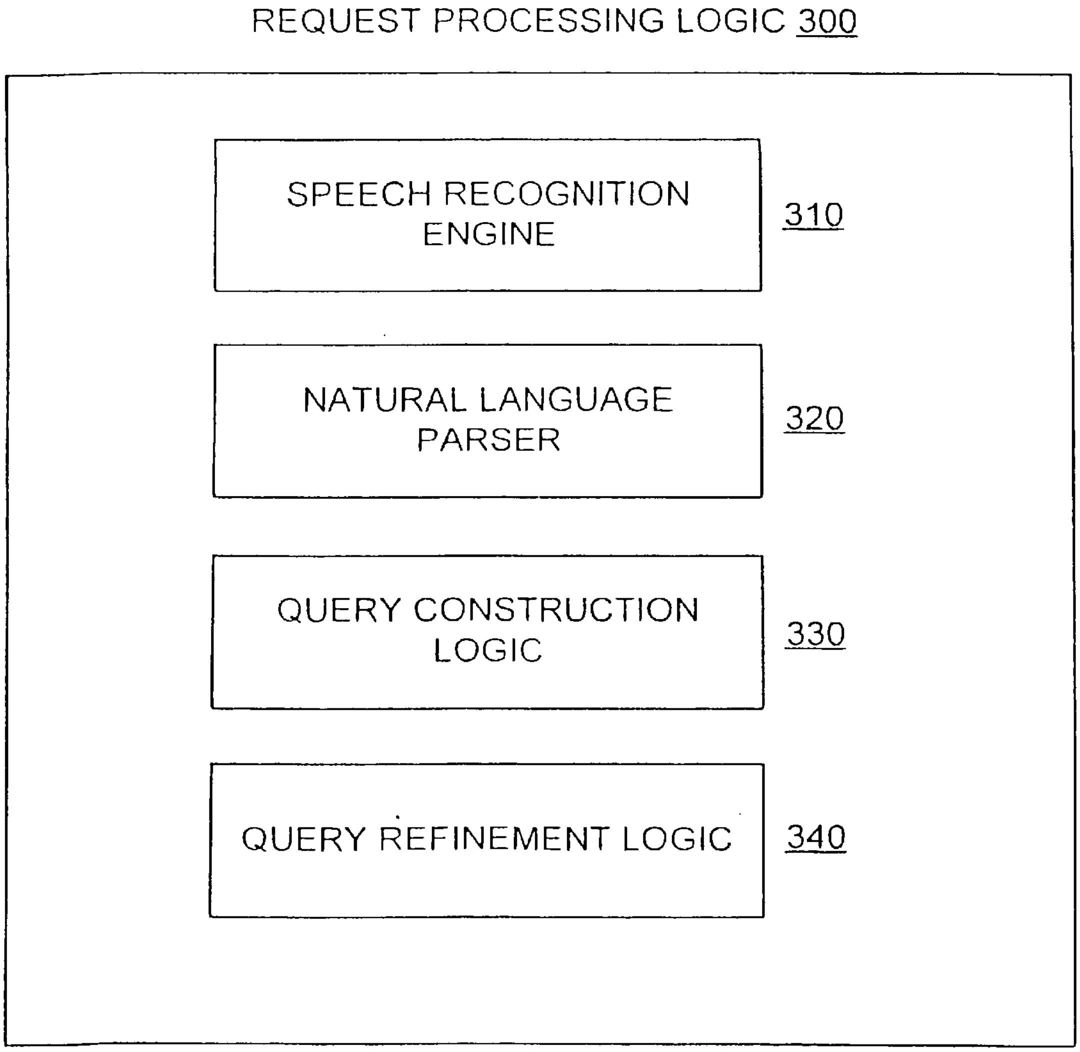

REQUEST PROCESSING LOGIC 300
SPEECH RECOGNITION ENGINE
310
NATURAL LANGUAGE PARSER
320
QUERY CONSTRUCTION LOGIC
330
QUERY REFINEMENT LOGIC
340

Fig. 3

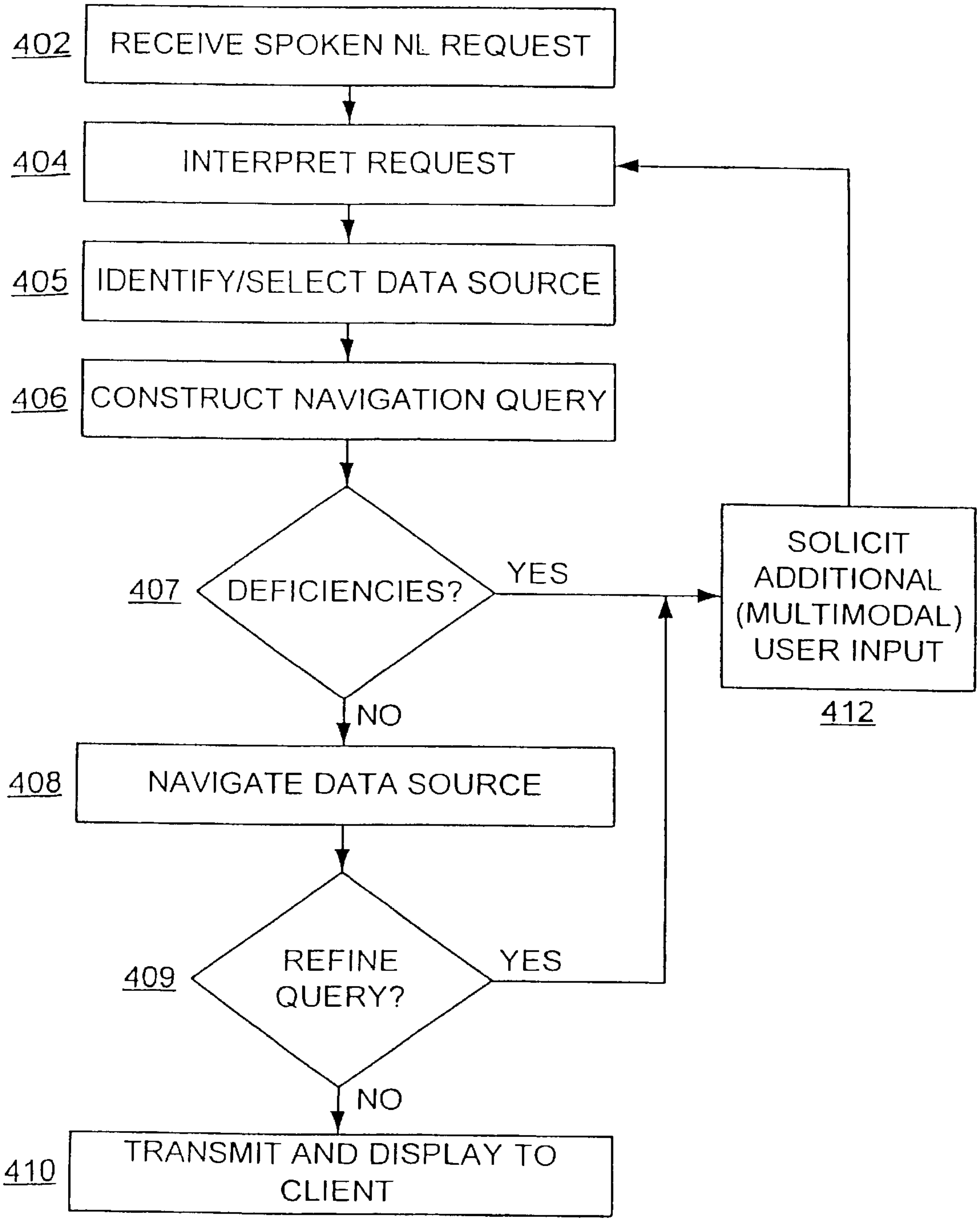

402
RECEIVE SPOKEN NL REQUEST
404
INTERPRET REQUEST
405
IDENTIFY/SELECT DATA SOURCE
406
CONSTRUCT NAVIGATION QUERY
407
DEFICIENCIES?
YES
NO
SOLICIT ADDITIONAL (MULTIMODAL) USER INPUT
412
408
NAVIGATE DATA SOURCE
409
REFINE QUERY?
YES
NO
410
TRANSMIT AND DISPLAY TO CLIENT

Fig. 4

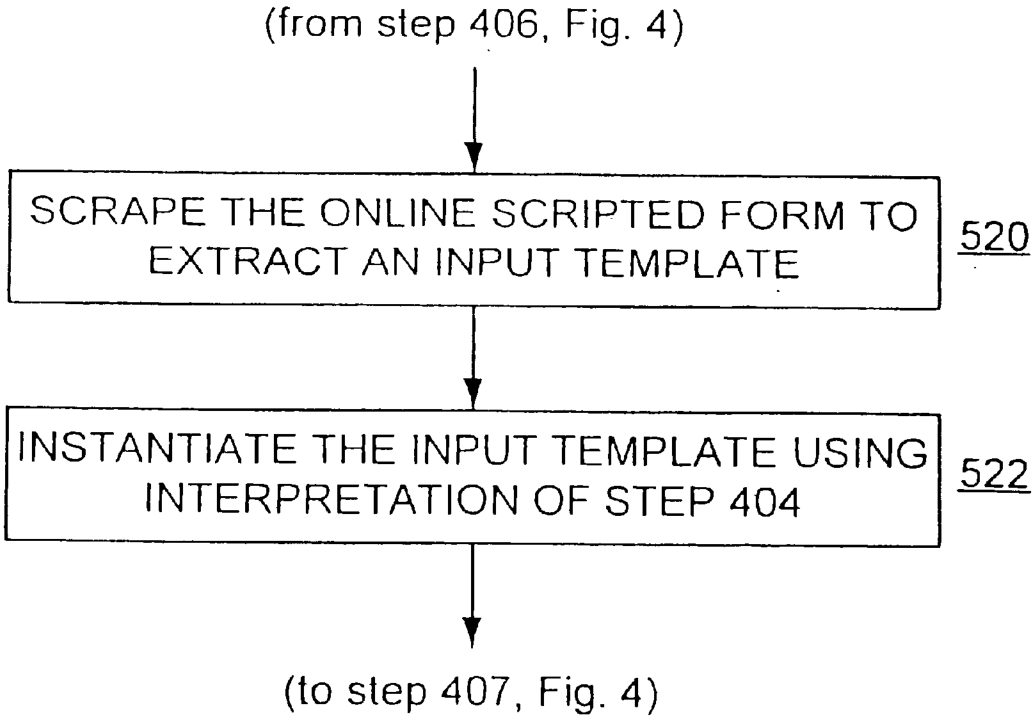

(from step 406, Fig. 4)
SCRAPE THE ONLINE SCRIPTED FORM TO EXTRACT AN INPUT TEMPLATE
520
INSTANTIATE THE INPUT TEMPLATE USING INTERPRETATION OF STEP 404
522
(to step 407, Fig. 4)

Fig. 5

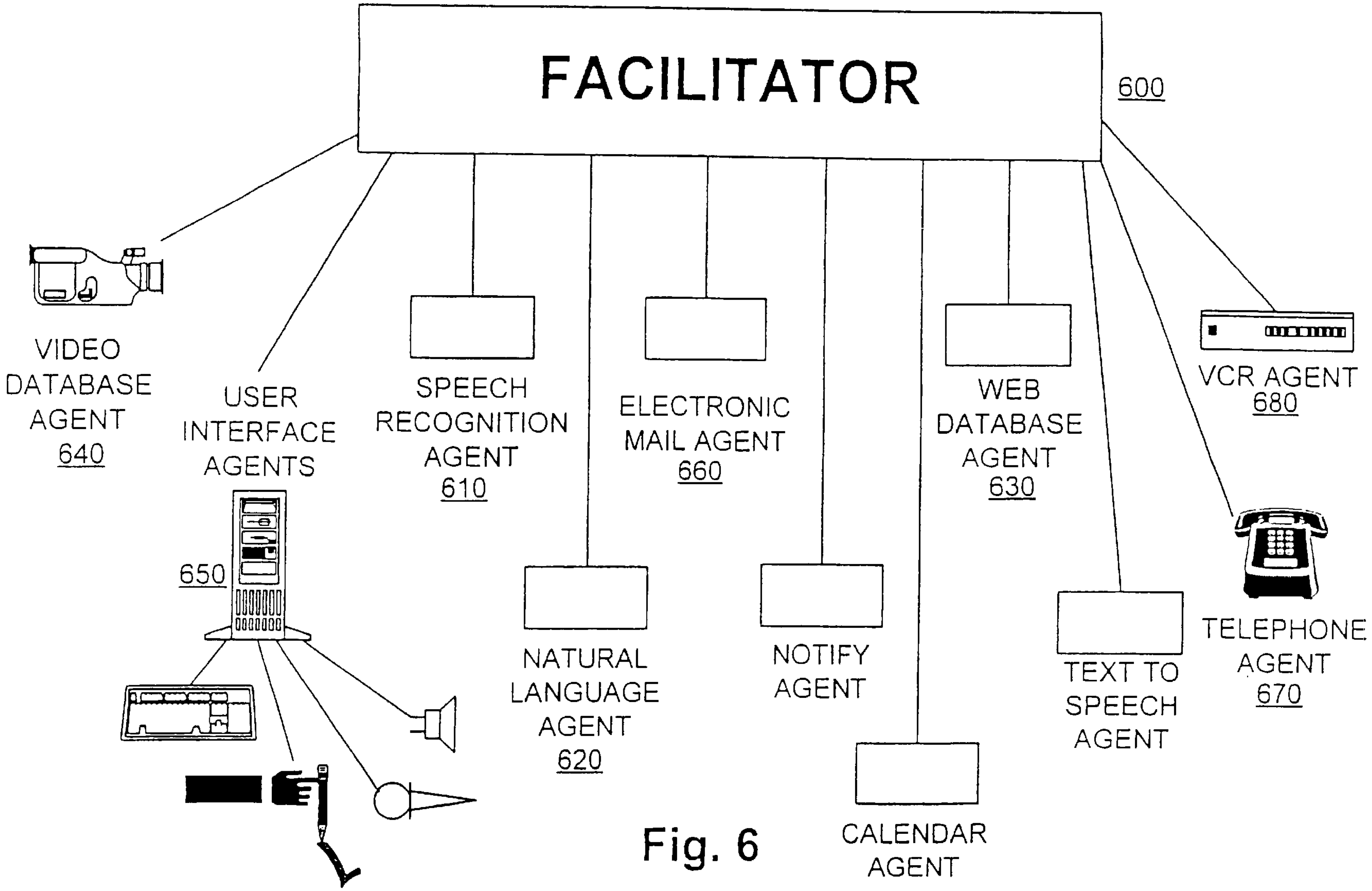

FACILITATOR
600
VIDEO DATABASE AGENT
640
USER INTERFACE AGENTS
650
SPEECH RECOGNITION AGENT
610
NATURAL LANGUAGE AGENT
620
ELECTRONIC MAIL AGENT
660
NOTIFY AGENT
CALENDAR AGENT
WEB DATABASE AGENT
630
TEXT TO SPEECH AGENT
TELEPHONE AGENT
670
VCR AGENT
680

Fig. 6

# MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

This application is a continuation of an application entitled NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK which was filed on Mar. 13, 2000 under Ser. No. 09/524,095 and which is a Continuation In Part of co-pending U.S. patent application Ser. No. 09/225,198, filed Jan. 5, 1999, Provisional U.S. patent application Ser. No. 60/124,718, filed Mar. 17, 1999, Provisional U.S. patent application Ser. No. 60/124,720, filed Mar. 17, 1999, and Provisional U.S. patent application Ser. No. 60/124,719, filed Mar. 17, 1999, from which applications priority is claimed and these application are incorporated herein by reference.

## BACKGROUND OF THE INVENTION

The present invention relates generally to the navigation of electronic data by means of spoken natural language requests, and to feedback mechanisms and methods for resolving the errors and ambiguities that may be associated with such requests.

As global electronic connectivity continues to grow, and the universe of electronic data potentially available to users continues to expand, there is a growing need for information navigation technology that allows relatively naive users to navigate and access desired data by means of natural language input. In many of the most important markets—including the home entertainment arena, as well as mobile computing—spoken natural language input is highly desirable, if not ideal. As just one example, the proliferation of high-bandwidth communications infrastructure for the home entertainment market (cable, satellite, broadband) enables delivery of movies-on-demand and other interactive multimedia content to the consumer's home television set. For users to take full advantage of this content stream ultimately requires interactive navigation of content databases in a manner that is too complex for user-friendly selection by means of a traditional remote-control clicker. Allowing spoken natural language requests as the input modality for rapidly searching and accessing desired content is an important objective for a successful consumer entertainment product in a context offering a dizzying range of database content choices. As further examples, this same need to drive navigation of (and transaction with) relatively complex data warehouses using spoken natural language requests applies equally to surfing the Internet/Web or other networks for general information, multimedia content, or e-commerce transactions.

In general, the existing navigational systems for browsing electronic databases and data warehouses (search engines, menus, etc.), have been designed without navigation via spoken natural language as a specific goal. So today's world is full of existing electronic data navigation systems that do not assume browsing via natural spoken commands, but rather assume text and mouse-click inputs (or in the case of TV remote controls, even less). Simply recognizing voice commands within an extremely limited vocabulary and grammar—the spoken equivalent of button/click input (e.g., speaking "channel 5" selects TV channel 5)—is really not sufficient by itself to satisfy the objectives described above. In order to deliver a true "win" for users, the voice-driven front-end must accept spoken natural language input in a manner that is intuitive to users. For example, the front-end should not require learning a highly specialized command language or format. More fundamentally, the front-end must allow users to speak directly in terms of what the user ultimately wants —e.g., "I'd like to see a Western film directed by Clint Eastwood" —as opposed to speaking in terms of arbitrary navigation structures (e.g., hierarchical layers of menus, commands, etc.) that are essentially artifacts reflecting constraints of the pre-existing text/click navigation system. At the same time, the front-end must recognize and accommodate the reality that a stream of naive spoken natural language input will, over time, typically present a variety of errors and/or ambiguities: e.g., garbled/unrecognized words (did the user say "Eastwood" or "Easter"?) and under-constrained requests ("Show me the Clint Eastwood movie"). An approach is needed for handling and resolving such errors and ambiguities in a rapid, user-friendly, non-frustrating manner.

What is needed is a methodology and apparatus for rapidly constructing a voice-driven front-end atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the step-by-step browsing architecture of the existing navigation system, and wherein any errors or ambiguities in user input are rapidly and conveniently resolved. The solution to this need should be compatible with the constraints of a multi-user, distributed environment such as the Internet/Web or a proprietary high-bandwidth content delivery network; a solution contemplating one-at-a-time user interactions at a single location is insufficient, for example.

## SUMMARY OF THE INVENTION

The present invention addresses the above needs by providing a system, method, and article of manufacture for mobile navigation of network-based electronic data sources in response to spoken input requests. When a spoken input request is received from a user using a mobile information appliance that communicates with a network server via an at least partially wireless communications system, it is interpreted, such as by using a speech recognition engine to extract speech data from acoustic voice signals, and using a language parser to linguistically parse the speech data. The interpretation of the spoken request can be performed on a computing device locally with the user, such as the mobile information appliance, or remotely from the user. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is then transmitted to a client device of the user. If the network data source is a database, the navigation query is constructed in the format of a database query language.

Typically, errors or ambiguities emerge in the interpretation of the spoken request, such that the system cannot instantiate a complete, valid navigational template. This is to be expected occasionally, and one preferred aspect of the invention is the ability to handle such errors and ambiguities in relatively graceful and user-friendly manner. Instead of simply rejecting such input and defaulting to traditional input modes or simply asking the user to try again, a preferred embodiment of the present invention seeks to converge rapidly toward instantiation of a valid navigational template by soliciting additional clarification from the user as necessary, either before or after a navigation of the data source, via multimodal input, i.e., by means of menu selection or other input modalities including and in addition to spoken input. This clarifying, multi-modal dialogue takes advantage of whatever partial navigational information has been gleaned from the initial interpretation of the user's spoken request. This clarification process continues until the system converges toward an adequately instantiated navigational template, which is in turn used to navigate the network-based data and retrieve the user's desired information. The retrieved information is transmitted across the network and presented to the user on a suitable client display device.

In a further aspect of the present invention, the construction of the navigation query includes extracting an input template for an online scripted interface to the data source and using the input template to construct the navigation query. The extraction of the input template can include dynamically scraping the online scripted interface.

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention, together with further advantages thereof, may best be understood by reference to the following description taken in conjunction with the accompanying drawings in which:

FIG. **1***a* illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with server-side processing of requests;

FIG. **1***b* illustrates another system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with client-side processing of requests;

FIG. **2** illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention for a mobile computing scenario;

FIG. **3** illustrates the functional logic components of a request processing module in accordance with an embodiment of the present invention;

FIG. **4** illustrates a process utilizing spoken natural language for navigating an electronic database in accordance with one embodiment of the present invention;

FIG. **5** illustrates a process for constructing a navigational query for accessing an online data source via an interactive, scripted (e.g., CGI) form; and

FIG. **6** illustrates an embodiment of the present invention utilizing a community of distributed, collaborating electronic agents.

## DETAILED DESCRIPTION OF THE INVENTION

### 1. System Architecture

#### a. Server-End Processing of Spoken Input

FIG. **1***a* is an illustration of a data navigation system driven by spoken natural language input, in accordance with one embodiment of the present invention. As shown, a user's voice input data is captured by a voice input device **102**, such as a microphone. Preferably voice input device **102** includes a button or the like that can be pressed or held-down to activate a listening mode, so that the system need not continually pay attention to, or be confused by, irrelevant background noise. In one preferred embodiment well-suited for the home entertainment setting, voice input device **102** is a portable remote control device with an integrated microphone, and the voice data is transmitted from device **102** preferably via infrared (or other wireless) link to communications box **104** (e.g., a set-top box or a similar communications device that is capable of retransmitting the raw voice data and/or processing the voice data) local to the user's environment and coupled to communications network **106**. The voice data is then transmitted across network **106** to a remote server or servers **108**. The voice data may preferably be transmitted in compressed digitized form, or alternatively—particularly where bandwidth constraints are significant—in analog format (e.g., via frequency modulated transmission), in the latter case being digitized upon arrival at remote server **108**.

At remote server **108**, the voice data is processed by request processing logic **300** in order to understand the user's request and construct an appropriate query or request for navigation of remote data source **110**, in accordance with the interpretation process exemplified in FIG. **4** and FIG. **5** and discussed in greater detail below. For purposes of executing this process, request processing logic **300** comprises functional modules including speech recognition engine **310**, natural language (NL) parser **320**, query construction logic **330**, and query refinement logic **340**, as shown in FIG. **3**. Data source **110** may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably resides on a central server or servers—which may or may not be the same as server **108**, depending on the storage and bandwidth needs of the application and the resources available to the practitioner. Data source **110** may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source **110** are navigated—i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user—using the processes of FIGS. **4** and **5** as described in greater detail below.

Once the desired information has been retrieved from data source **110**, it is electronically transmitted via network **106** to the user for viewing on client display device **112**. In a preferred embodiment well-suited for the home entertainment setting, display device **112** is a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device **112** is coupled to or integrated with a communications box (which is preferably the same as communications box **104**, but may also be a separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network **106**.

Network **106** is a two-way electronic communications network and may be embodied in electronic communication infrastructure including coaxial (cable television) lines, DSL, fiber-optic cable, traditional copper wire (twisted pair), or any other type of hardwired connection. Network **106** may also include a wireless connection such as a satellite-based connection, cellular connection, or other type of wireless connection. Network **106** may be part of the Internet and may support TCP/IP communications, or may be embodied in a proprietary network, or in any other electronic communications network infrastructure, whether packet-switched or connection-oriented. A design consideration is that network **106** preferably provide suitable bandwidth depending upon the nature of the content anticipated for the desired application.

#### b. Client-End Processing of Spoken Input

FIG. **1***b* is an illustration of a data navigation system driven by spoken natural language input, in accordance with a second embodiment of the present invention. Again, a user's voice input data is captured by a voice input device

**102**, such as a microphone. In the embodiment shown in FIG. **1***b*, the voice data is transmitted from device **202** to requests processing logic **300**, hosted on a local speech processor, for processing and interpretation. In the preferred embodiment illustrated in FIG. **1***b*, the local speech processor is conveniently integrated as part of communications box **104**, although implementation in a physically separate (but communicatively coupled) unit is also possible as will be readily apparent to those of skill in the art. The voice data is processed by the components of request processing logic **300** in order to understand the user's request and construct an appropriate query or request for navigation of remote data source **110**, in accordance with the interpretation process exemplified in FIGS. **4** and **5** as discussed in greater detail below.

The resulting navigational query is then transmitted electronically across network **106** to data source **110**, which preferably resides on a central server or servers **108**. As in FIG. **1***a*, data source **110** may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source **110** are then navigated—i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user—preferably using the process of FIGS. **4** and **5** as described in greater detail below. Once the desired information has been retrieved from data source **110**, it is electronically transmitted via network **106** to the user for viewing on client display device **112**.

In one embodiment in accordance with FIG. **1***b* and well-suited for the home entertainment setting, voice input device **102** is a portable remote control device with an integrated microphone, and the voice data is transmitted from device **102** preferably via infrared (or other wireless) link to the local speech processor. The local speech processor is coupled to communications network **106**, and also preferably to client display device **112** (especially for purposes of query refinement transmissions, as discussed below in connection with FIG. **4**, step **412**), and preferably may be integrated within or coupled to communications box **104**. In addition, especially for purposes of a home entertainment application, display device **112** is preferably a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device **112** is coupled to a communications box (which is preferably the same as communications box **104**, but may also be a physically separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network **106**.

Design considerations favoring server-side processing and interpretation of spoken input requests, as exemplified in FIG. **1***a*, include minimizing the need to distribute costly computational hardware and software to all client users in order to perform speech and language processing. Design considerations favoring client-side processing, as exemplified in FIG. **1***b*, include minimizing the quantity of data sent upstream across the network from each client, as the speech recognition is performed before transmission across the network and only the query data and/or request needs to be sent, thus reducing the upstream bandwidth requirements.

c. Mobile Client Embodiment

A mobile computing embodiment of the present invention may be implemented by practitioners as a variation on the embodiments of either FIG. **1***a* or FIG. **1***b*. For example, as depicted in FIG. **2**, a mobile variation in accordance with the server-side processing architecture illustrated in FIG. **1***a* may be implemented by replacing voice input device **102**, communications box **104**, and client display device **112**, with an integrated, mobile, information appliance **202** such as a cellular telephone or wireless personal digital assistant (wireless PDA). Mobile information appliance **202** essentially performs the functions of the replaced components. Thus, mobile information appliance **202** receives spoken natural language input requests from the user in the form of voice data, and transmits that data (preferably via wireless data receiving station **204**) across communications network **206** for server-side interpretation of the request, in similar fashion as described above in connection with FIG. **1**. Navigation of data source **210** and retrieval of desired information likewise proceeds in an analogous manner as described above. Display information transmitted electronically back to the user across network **206** is displayed for the user on the display of information appliance **202**, and audio information is output through the appliance's speakers.

Practitioners will further appreciate, in light of the above teachings, that if mobile information appliance **202** is equipped with sufficient computational processing power, then a mobile variation of the client-side architecture exemplified in FIG. **2** may similarly be implemented. In that case, the modules corresponding to request processing logic **300** would be embodied locally in the computational resources of mobile information appliance **202**, and the logical flow of data would otherwise follow in a manner analogous to that previously described in connection with FIG. **1***b*.

As illustrated in FIG. **2**, multiple users, each having their own client input device, may issue requests, simultaneously or otherwise, for navigation of data source **210**. This is equally true (though not explicitly drawn) for the embodiments depicted in FIGS. **1***a* and **1***b*. Data source **210** (or **100**), being a network accessible information resource, has typically already been constructed to support access requests from simultaneous multiple network users, as known by practitioners of ordinary skill in the art. In the case of server-side speech processing, as exemplified in FIGS. **1***a* and **2**, the interpretation logic and error correction logic modules are also preferably designed and implemented to support queuing and multi-tasking of requests from multiple simultaneous network users, as will be appreciated by those of skill in the art.

It will be apparent to those skilled in the art that additional implementations, permutations and combinations of the embodiments set forth in FIGS. **1***a*, **1***b*, and **2** may be created without straying from the scope and spirit of the present invention. For example, practitioners will understand, in light of the above teachings and design considerations, that it is possible to divide and allocate the functional components of request processing logic **300** between client and server. For example, speech recognition—in entirety, or perhaps just early stages such as feature extraction—might be performed locally on the client end, perhaps to reduce bandwidth requirements, while natural language parsing and other necessary processing might be performed upstream on the server end, so that more extensive computational power need not be distributed locally to each client. In that case, corresponding portions of request processing logic **300**, such as speech recognition engine **310** or portions thereof, would reside locally at the client as in FIG. **1***b*, while other component modules would be hosted at the server end as in FIGS. **1***a* and **2**.

Further, practitioners may choose to implement the each of the various embodiments described above on any number of different hardware and software computing platforms and environments and various combinations thereof, including, by way of just a few examples: a general-purpose hardware microprocessor such as the Intel Pentium series; operating system software such as Microsoft Windows/CE, Palm OS, or Apple Mac OS (particularly for client devices and client-side processing), or Unix, Linux, or Windows/NT (the latter three particularly for network data servers and server-side processing), and/or proprietary information access platforms such as Microsoft's WebTV or the Diva Systems video-on-demand system.

2. Processing Methodology

The present invention provides a spoken natural language interface for interrogation of remote electronic databases and retrieval of desired information. A preferred embodiment of the present invention utilizes the basic methodology outlined in the flow diagram of FIG. **4** in order to provide this interface. This methodology will now be discussed.

a. Interpreting Spoken Natural Language Requests

At step **402**, the user's spoken request for information is initially received in the form of raw (acoustic) voice data by a suitable input device, as previously discussed in connection with FIGS. **1–2**. At step **404** the voice data received from the user is interpreted in order to understand the user's request for information. Preferably this step includes performing speech recognition in order to extract words from the voice data, and further includes natural language parsing of those words in order to generate a structured linguistic representation of the user's request.

Speech recognition in step **404** is performed using speech recognition engine **310**. A variety of commercial quality, speech recognition engines are readily available on the market, as practitioners will know. For example, Nuance Communications offers a suite of speech recognition engines, including Nuance 6, its current flagship product, and Nuance Express, a lower cost package for entry-level applications. As one other example, IBM offers the ViaVoice speech recognition engine, including a low-cost shrink-wrapped version available through popular consumer distribution channels. Basically, a speech recognition engine processes acoustic voice data and attempts to generate a text stream of recognized words.

Typically, the speech recognition engine is provided with a vocabulary lexicon of likely words or phrases that the recognition engine can match against its analysis of acoustical signals, for purposes of a given application. Preferably, the lexicon is dynamically adjusted to reflect the current user context, as established by the preceding user inputs. For example, if a user is engaged in a dialogue with the system about movie selection, the recognition engine's vocabulary may preferably be adjusted to favor relevant words and phrases, such as a stored list of proper names for popular movie actors and directors, etc. Whereas if the current dialogue involves selection and viewing of a sports event, the engine's vocabulary might preferably be adjusted to favor a stored list of proper names for professional sports teams, etc. In addition, a speech recognition engine is provided with language models that help the engine predict the most likely interpretation of a given segment of acoustical voice data, in the current context of phonemes or words in which the segment appears. In addition, speech recognition engines often echo to the user, in more or less real-time, a transcription of the engine's best guess at what the user has said, giving the user an opportunity to confirm or reject.

In a further aspect of step **404**, natural language interpreter (or parser) **320** linguistically parses and interprets the textual output of the speech recognition engine. In a preferred embodiment of the present invention, the natural-language interpreter attempts to determine both the meaning of spoken words (semantic processing) as well as the grammar of the statement (syntactic processing), such as the Gemini Natural Language Understanding System developed by SRI International. The Gemini system is described in detail in publications entitled "Gemini: A Natural Language System for Spoken-Language Understanding" and "Interleaving Syntax and Semantics in an Efficient Bottom-Up Parser," both of which are currently available online at http://www.ai.sri.com/natural-language/projects/arpa-sls/nat-lang.html. (Copies of those publications are also included in an information disclosure statement submitted herewith, and are incorporated herein by this reference). Briefly, Gemini applies a set of syntactic and semantic grammar rules to a word string using a bottom-up parser to generate a logical form, which is a structured representation of the context-independent meaning of the string. Gemini can be used with a variety of grammars, including general English grammar as well as application-specific grammars. The Gemini parser is based on "unification grammar," meaning that grammatical categories incorporate features that can be assigned values; so that when grammatical category expressions are matched in the course of parsing or semantic interpretation, the information contained in the features is combined, and if the feature values are incompatible the match fails.

It is possible for some applications to achieve a significant reduction in speech recognition error by using the natural-language processing system to re-score recognition hypotheses. For example, the grammars defined for a language parser like Gemini may be compiled into context-free grammar that, in turn, can be used directly as language models for speech recognition engines like the Nuance recognizer. Further details on this methodology are provided in the publication "Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS" which is currently available online through http://www.ai.sri.com/natural-language/projects/arpa-sls/spnl-int.html. A copy of this publication is included in an information disclosure submitted herewith, and is incorporated herein by this reference.

In an embodiment of the present invention that may be preferable for some applications, the natural language interpreter "learns" from the past usage patterns of a particular user or of groups of users. In such an embodiment, the successfully interpreted requests of users are stored, and can then be used to enhance accuracy by comparing a current request to the stored requests, thereby allowing selection of a most probable result.

b. Constructing Navigation Queries

In step **405** request processing logic **300** identifies and selects an appropriate online data source where the desired information (in this case, current weather reports for a given city) can be found. Such selection may involve look-up in a locally stored table, or possibly dynamic searching through an online search engine, or other online search techniques. For some applications, an embodiment of the present invention may be implemented in which only access to a particular data source (such as a particular vendor's proprietary content database) is supported; in that case, step **405** may be trivial or may be eliminated entirely.

Step **406** attempts to construct a navigation query, reflecting the interpretation of step **404**. This operation is preferably performed by query construction logic **330**.

A "navigation query" means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information. In other words, a navigation query is constructed such that it includes whatever content and structure is required in order to access desired information electronically from a particular database or data source of interest.

For example, for many existing electronic databases, a navigation query can be embodied using a formal database query language such as Standard Query Language (SQL). For many databases, a navigation query can be constructed through a more user-friendly interactive front-end, such as a series of menus and/or interactive forms to be selected or filled in. SQL is a standard interactive and programming language for getting information from and updating a database. SQL is both an ANSI and an ISO standard. As is well known to practitioners, a Relational Database Management System (RDBMS), such as Microsoft's Access, Oracle's Oracle**7**, and Computer Associates' CA-OpenIngres, allow programmers to create, update, and administer a relational database. Practitioners of ordinary skill in the art will be thoroughly familiar with the notion of database navigation through structured query, and will be readily able to appreciate and utilize the existing data structures and navigational mechanisms for a given database, or to create such structures and mechanisms where desired.

In accordance with the present invention, the query constructed in step **406** must reflect the user's request as interpreted by the speech recognition engine and the NL parser in step **404**. In embodiments of the present invention wherein data source **110** (or **210** in the corresponding embodiment of FIG. **2**) is a structured relational database or the like, step **406** of the present invention may entail constructing an appropriate Structured Query Language (SQL) query or the like, or automatically filling out a front-end query form, series of menus or the like, as described above.

In many existing Internet (and Intranet) applications, an online electronic data source is accessible to users only through the medium of interaction with a so-called Common Gateway Interface (CGI) script. Typically the user who visits a web site of this nature must fill in the fields of an online interactive form. The online form is in turn linked to a CGI script, which transparently handles actual navigation of the associated data source and produces output for viewing by the user's web browser. In other words, direct user access to the data source is not supported, only mediated access through the form and CGI script is offered.

For applications of this nature, an advantageous embodiment of the present invention "scrapes" the scripted online site where information desired by a user may be found in order to facilitate construction of an effective navigation query. For example, suppose that a user's spoken natural language request is: "What's the weather in Miami?" After this request is received at step **402** and interpreted at step **404**, assume that step **405** determines that the desired weather information is available online through the medium of a CGI-scripted interactive form. Step **406** is then preferably carried out using the expanded process diagrammed in FIG. **5**. In particular, at sub-step **520**, query construction logic **330** electronically "scrapes" the online interactive form, meaning that query construction logic **330** automatically extracts the format and structure of input fields accepted by the online form. At sub-step **522**, a navigation query is then constructed by instantiating (filling in) the extracted input format—essentially an electronic template—in a manner reflecting the user's request for information as interpreted in step **404**. The flow of control then returns to step **407** of FIG. **4**. Ultimately, when the query thus constructed by scraping is used to navigate the online data source in step **408**, the query effectively initiates the same scripted response as if a human user had visited the online site and had typed appropriate entries into the input fields of the online form.

In the embodiment just described, scraping step **520** is preferably carried out with the assistance of an online extraction utility such as WebL. WebL is a scripting language for automating tasks on the World Wide Web. It is an imperative, interpreted language that has built-in support for common web protocols like HTTP and FTP, and popular data types like HTML and XML. WebL's implementation language is Java, and the complete source code is available from Compaq. In addition, step **520** is preferably performed dynamically when necessary—in other words, on-the-fly in response to a particular user query—but in some applications it may be possible to scrape relatively stable (unchanging) web sites of likely interest in advance and to cache the resulting template information.

It will be apparent, in light of the above teachings, that preferred embodiments of the present invention can provide a spoken natural language interface atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the linear browsing architecture or other artifacts of an existing menu/text/click navigation system. For example, users of an appropriate embodiment of the present invention for a video-on-demand application can directly speak the natural request: "Show me the movie 'Unforgiven'"—instead of walking step-by-step through a typically linear sequence of genre/title/actor/director menus, scrolling and selecting from potentially long lists on each menu, or instead of being forced to use an alphanumeric keyboard that cannot be as comfortable to hold or use as a lightweight remote control. Similarly, users of an appropriate embodiment of the present invention for a web-surfing application in accordance with the process shown in FIG. **5** can directly speak the natural request: "Show me a one-month price chart for Microsoft stock" —instead of potentially having to navigate to an appropriate web site, search for the right ticker symbol, enter/select the symbol, and specify display of the desired one-month price chart, each of those steps potentially involving manual navigation and data entry to one or more different interaction screens. (Note that these examples are offered to illustrate some of the potential benefits offered by appropriate embodiments of the present invention, and not to limit the scope of the invention in any respect.)

c. Error Correction

Several problems can arise when attempting to perform searches based on spoken natural language input. As indicated at decision step **407** in the process of FIG. **4**, certain deficiencies may be identified during the process of query construction, before search of the data source is even attempted. For example, the user's request may fail to specify enough information in order to construct a navigation query that is specific enough to obtain a satisfactory search result. For example, a user might orally request "what's the weather?" whereas the national online data source identified in step **405** and scraped in step **520** might require specifying a particular city.

Additionally, certain deficiencies and problems may arise following the navigational search of the data source at step **408**, as indicated at decision step **409** in FIG. **4**. For example, with reference to a video-on-demand application, a user may wish to see the movie "Unforgiven", but perhaps the user can't recall name of the film, but knows it was directed by and starred actor Clint Eastwood. A typical video-on-demand database might indeed be expected to allow queries specifying the name of a leading actor and/or director, but in the case of this query—as in many cases—that will not be enough to narrow the search to a single film, and additional user input in some form is required.

In the event that one or more deficiencies in the user's spoken request, as processed, result in the problems described, either at step **407** or **409**, some form of error handling is in order. A straightforward, crude technique might be for the system to respond simply "input not understood/insufficient, please try again." However, that approach will likely result in frustrated users, and is not optimal or even acceptable for most applications. Instead, a preferred technique in accordance with the present invention handles such errors and deficiencies in user input at step **412**, whether detected at step **407** or step **409**, by soliciting additional input from the user in a manner taking advantage of the partial construction already performed and via user interface modalities in addition to spoken natural language ("multi-modality"). This supplemental interaction is preferably conducted through client display device **112** (**202**, in the embodiment of FIG. **2**), and may include textual, graphical, audio and/or video media. Further details and examples are provided below. Query refinement logic **340** preferably carries out step **412**. The additional input received from the user is fed into and augments interpreting step **404**, and query construction step **406** is likewise repeated with the benefit of the augmented interpretation. These operations, and subsequent navigation step **408**, are preferably repeated until no remaining problems or deficiencies are identified at decision points **407** or **409**. Further details and examples for this query refinement process are provided immediately below.

Consider again the example in which the user of a video-on-demand application wishes to see "Unforgiven" but can only recall that it was directed by and starred Clint Eastwood. First, it bears noting that using a prior art navigational interface, such as a conventional menu interface, will likely be relatively tedious in this case. The user can proceed through a sequence of menus, such as Genre (select "western"), Title (skip), Actor ("Clint Eastwood"), and Director ("Clint Eastwood"). In each case—especially for the last two items—the user would typically scroll and select from fairly long lists in order to enter his or her desired name, or perhaps use a relatively couch-unfriendly keypad to manually type the actor's name twice.

Using a preferred embodiment of the present invention, the user instead speaks aloud, holding remote control microphone **102**, "I want to see that movie starring and directed by Clint Eastwood. Can't remember the title." At step **402** the voice data is received. At step **404** the voice data is interpreted. At step **405** an appropriate online data source is selected (or perhaps the system is directly connected to a proprietary video-on-demand provider). At step **406** a query is automatically constructed by the query construction logic **330** specifying "Clint Eastwood" in both the actor and director fields. Step **407** detects no obvious problems, and so the query is electronically submitted and the data source is navigated at step **408**, yielding a list of several records satisfying the query (e.g., "Unforgiven", "True Crime", "Absolute Power", etc.). Step **409** detects that additional user input is needed to further refine the query in order to select a particular film for viewing.

At that point, in step **412** query refinement logic **340** might preferably generate a display for client display device **112** showing the (relatively short) list of film titles that satisfy the user's stated constraints. The user can then preferably use a relatively convenient input modality, such as buttons on the remote control, to select the desired title from the menu. In a further preferred embodiment, the first title on the list is highlighted by default, so that the user can simply press an "OK" button to choose that selection. In a further preferred feature, the user can mix input modalities by speaking a response like "I want number one on the list." Alternatively, the user can preferably say, "Let's see Unforgiven," having now been reminded of the title by the menu display.

Utilizing the user's supplemental input, request processing logic **300** iterates again through steps **404** and **406**, this time constructing a fully-specified query that specifically requests the Eastwood film "Unforgiven." Step **408** navigates the data source using that query and retrieves the desired film, which is then electronically transmitted in step **410** from network server **108** to client display device **112** via communications network **106**.

Now consider again the example in which the user of a web surfing application wants to know his or her local weather, and simply asks, "what's the weather?" At step **402** the voice data is received. At step **404** the voice data is interpreted. At step **405** an online web site providing current weather information for major cities around the world is selected. At step **406** and sub-step **520**, the online site is scraped using a WebL-style tool to extract an input template for interacting with the site. At sub-step **522**, query construction logic **330** attempts to construct a navigation query by instantiating the input template, but determines (quite rightly) that a required field—name of city—cannot be determined from the user's spoken request as interpreted in step **404**. Step **407** detects this deficiency, and in step **412** query refinement logic **340** preferably generates output for client display device **112** soliciting the necessary supplemental input. In a preferred embodiment, the output might display the name of the city where the user is located highlighted by default. The user can then simply press an "OK" button—or perhaps mix modalities by saying "yes, exactly" —to choose that selection. A preferred embodiment would further display an alphabetical scrollable menu listing other major cities, and/or invite the user to speak or select the name of the desired city.

Here again, utilizing the user's supplemental input, request processing logic **300** iterates through steps **404** and **406**. This time, in performing sub-step **520**, a cached version of the input template already scraped in the previous iteration might preferably be retrieved. In sub-step **522**, query construction logic **330** succeeds this time in instantiating the input template and constructing an effective query, since the desired city has now been clarified. Step **408** navigates the data source using that query and retrieves the desired weather information, which is then electronically transmitted in step **410** from network server **108** to client display device **112** via communications network **106**.

It is worth noting that in some instances, there may be details that are not explicitly provided by the user, but that query construction logic **330** or query refinement logic **340** may preferably deduce on their own through reasonable assumptions, rather than requiring the use to provide explicit clarification. For example, in the example previously described regarding a request for a weather report, in some applications it might be preferable for the system to simply assume that the user means a weather report for his or her home area and to retrieve that information, if the cost of doing so is not significantly greater than the cost of asking the user to clarify the query. Making such an assumption might be even more strongly justified in a preferred embodiment, as described earlier, where user histories are tracked, and where such history indicates that a particular user or group of users typically expect local information when asking for a weather forecast. At any rate, in the event such an assumption is made, if the user actually intended to request the weather for a different city, the user would then need to ask his or her question again. It will be apparent to practitioners, in light of the above teachings, that the choice of whether to program query construction logic **330** and query refinement logic **340** to make make particular assumptions will typically involve trade-offs involving user conveience that can be assessed in the context of specific applications.

3. Open Agent Architecture (OAA®)

Open Agent Architecture™ (OAA®) is a software platform, developed by the assignee of the present invention, that enables effective, dynamic collaboration among communities of distributed electronic agents. OAA is described in greater detail in co-pending U.S. patent application Ser. No. 09/225,198, which has been incorporated herein by reference. Very briefly, the functionality of each client agent is made available to the agent community through registration of the client agent's capabilities with a facilitator. A software "wrapper" essentially surrounds the underlying application program performing the services offered by each client. The common infrastructure for constructing agents is preferably supplied by an agent library. The agent library is preferably accessible in the runtime environment of several different programming languages. The agent library preferably minimizes the effort required to construct a new system and maximizes the ease with which legacy systems can be "wrapped" and made compatible with the agent-based architecture of the present invention. When invoked, a client agent makes a connection to a facilitator, which is known as its parent facilitator. Upon connection, an agent registers with its parent facilitator a specification of the capabilities and services it can provide, using a high-level, declarative Interagent Communication Language ("ICL") to express those capabilities. Tasks are presented to the facilitator in the form of ICL goal expressions. When a facilitator determines that the registered capabilities of one of its client agents will help satisfy a current goal or sub-goal thereof, the facilitator delegates that sub-goal to the client agent in the form of an ICL request. The client agent processes the request and returns answers or information to the facilitator. In processing a request, the client agent can use ICL to request services of other agents, or utilize other infrastructure services for collaborative work. The facilitator coordinates and integrates the results received from different client agents on various sub-goals, in order to satisfy the overall goal.

OAA provides a useful software platform for building systems that integrate spoken natural language as well as other user input modalities. For example, see the above-referenced co-pending patent application, especially FIG. **13** and the corresponding discussion of a "multi-modal maps" application, and FIG. **12** and the corresponding discussion of a "unified messaging" application. Another example is the InfoWiz interactive information kiosk developed by the assignee and described in the document entitled "InfoWiz: An Animated Voice Interactive Information System" available online at http://www.ai.sri.com/~oaa/applications.html. A copy of the InfoWhiz document is provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference. A further example is the "CommandTalk" application developed by the assignee for the U.S. military, as described online at http://www.ai.sri.com/~lesaf/commandtalk.html and in the following publications, copies of which are provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference:

"CommandTalk: A Spoken-Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, D.C., pp. 1–7, Association for Computational Linguistics

"The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty-Seventh Annual Meeting of the ACL", pp. 183–190, University of Maryland, College Park, Md., Association for Computational Linguistics

"Interpreting Language in Context in CommandTalk", 1999, by John Dowding and Elizabeth Owen Bratt and Sharon Goldwater, in "Communicative Agents: The Use of Natural Language in Embodied Systems", pp. 63–67, Association for Computing Machinery (ACM) Special Interest Group on Artificial Intelligence (SIGART), Seattle, Wash.

For some applications and systems, OAA can provide an advantageous platform for constructing embodiments of the present invention. For example, a representative application is now briefly presented, with reference to FIG. **6**. If the statement "show me movies starring John Wayne" is spoken into the voice input device, the voice data for this request will be sent by UI agent **650** to facilitator **600**, which in turn will ask natural language (NL) agent **620** and speech recognition agent **610** to interpret the query and return the interpretation in ICL format. The resulting ICL goal expression is then routed by the facilitator to appropriate agents—in this case, video-on-demand database agent **640**—to execute the request. Video database agent **640** preferably includes or is coupled to an appropriate embodiment of query construction logic **330** and query refinement logic **340**, and may also issue ICL requests to facilitator **600** for additional assistance—e.g., display of menus and capture of additional user input in the event that query refinement is needed—and facilitator **600** will delegate such requests to appropriate client agents in the community. When the desired video content is ultimately retrieved by video database agent **640**, UI agent **650** is invoked by facilitator **600** to display the movie.

Other spoken user requests, such as a request for the current weather in New York City or for a stock quote, would eventually lead facilitator to invoke web database agent **630** to access the desired information from an appropriate Internet site. Here again, web database agent **630** preferably includes or is coupled to an appropriate embodiment of query construction logic **330** and query refinement logic **340**, including a scraping utility such as WebL. Other spoken requests, such as a request to view recent emails or access voice mail, would lead the facilitator to invoke the appropriate email agent **660** and/or telephone agent **680**. A request to record a televised program of interest might lead facilitator **600** to invoke web database agent **630** to return televised program schedule information, and then invoke VCR controller agent **680** to program the associated VCR unit to record the desired television program at the scheduled time.

Control and connectivity embracing additional electronic home appliances (e.g., microwave oven, home surveillance system, etc.) can be integrated in comparable fashion. Indeed, an advantage of OAA-based embodiments of the present invention, that will be apparent to practitioners in light of the above teachings and in light of the teachings disclosed in the cited co-pending patent applications, is the relative ease and flexibility with which additional service agents can be plugged into the existing platform, immediately enabling the facilitator to respond dynamically to spoken natural language requests for the corresponding services.

4. Further Embodiments and Equivalents

While the present invention has been described in terms of several preferred embodiments, there are many alterations, permutations, and equivalents that may fall within the scope of this invention. It should also be noted that there are many alternative ways of implementing the methods and apparatuses of the present invention. It is therefore intended that the following appended claims be interpreted as including all such alterations, permutations, and equivalents as fall within the true spirit and scope of the present invention.

What is claimed is:

**1**. A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d) utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

**2**. The method of claim **1**, wherein the step of rendering the interpretation of the spoken request is performed by the mobile information appliance.

**3**. The method of claim **1**, wherein the step of rendering the interpretation of the spoken request is performed by the mobile information appliance.

**4**. The method of claim **1**, further comprising the steps of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source.

**5**. The method of claim **1**, wherein the data link includes a cellular telephone system.

**6**. The method of claim **1**, wherein steps (a)–(d) are performed with respect to multiple users.

**7**. The method of claim **1**, wherein the mobile information appliance is a wireless telephone.

**8**. The method of claim **1**, wherein the mobile information appliance is a portable computing device.

**9**. The method of claim **8**, wherein the portable computing device is a personal digital assistant.

**10**. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

**11**. The computer program of claim **10**, wherein the rendering of the interpretation of the spoken request is performed at the one or more network servers.

**12**. The computer program of claim **10**, wherein the rendering of the interpretation of the spoken request is performed by the mobile information appliance.

**13**. The computer program of claim **10**, further comprising a code segment that solicits additional input from the user, including user interaction in a modality different than the original request; a code segment that refines the navigation query, based upon the additional input; and a code segment that uses the refined navigation query to select a portion of the electronic data source.

**14**. The computer program of claim **10**, wherein the data link includes a wireless telephone system.

**15**. The computer program of claim **10**, wherein code segments (a)–(d) are executed with respect to multiple users.

**16**. The computer program of claim **10**, wherein the mobile information appliance is a wireless telephone.

**17**. The computer program of claim **10**, wherein the mobile information appliance is a portable computing device.

**18**. The computer program of claim **17**, wherein the portable computing device is a personal digital assistant.

**19**. A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

**20**. The system of claim **19**, wherein the spoken language processing logic renders the interpretation of the spoken request at the one or more network servers.

**21**. The system of claim **19**, wherein the spoken language processing logic renders the interpretation of the spoken request at the mobile information appliance.

**22**. The system of claim **19**, further comprising user interaction logic operable to solicit additional input from the user, including user interaction in a modality different than the original request; and query refining logic operable to refine the navigation query based upon the additional input; wherein the navigation logic users the refined navigation query to select a portion of the electronic data source.

**23**. The system of claim **19**, wherein the data link includes a cellular telephone system.

**24**. The system of claim **19**, wherein the system operates with respect to multiple users.

**25**. The system of claim **19**, wherein the mobile information appliance is a wireless telephone.

**26**. The system of claim **19**, wherein the mobile information appliance is a portable computing device.

**27**. The system of claim **26**, wherein the portable computing device is a personal digital assistant.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE PATENT TRIAL AND APPEAL BOARD

---

GOOGLE LLC
Petitioner

v.

IPA TECHNOLOGIES INC.
Patent Owner

---

Patent No. 6,757,718

---

# DECLARATION OF DR. DAN R. OLSEN JR.

## TABLE OF CONTENTS

I, Dr. Dan R. Olsen Jr., declare as follows:

## I. INTRODUCTION

1. I have been retained by Google LLC ("Petitioner") as an independent expert consultant in this proceeding before the United States Patent and Trademark Office ("PTO") regarding U.S. Patent No. 6,757,718 ("the '718 patent") (Ex. 1001). I have been asked to consider whether certain references disclose or suggest the features recited in claims 1-27 ("the challenged claims") of the '718 patent. My opinions are set forth below.

2. I am being compensated at my rate of $500 per hour for the time I spend on this matter. My compensation is in no way contingent on the nature of my findings, the presentation of my findings in testimony, or the outcome of this or any other proceeding. I have no other interest in this proceeding.

## II. BACKGROUND AND QUALIFICATIONS

3. I have more than 35 years of experience in computer science and human-computer interaction (HCI). I hold a doctorate in Computing and Information from the University of Pennsylvania. For 3 ½ years I was an Assistant Professor of Computer Science at Arizona State University. I then served for 30 years on the faculty of Brigham Young University, retiring as a full professor in 2015. During that time at BYU, I also served as the chair of the Department of

Computer Science. I took leave from BYU in 1996 to become the founding director of the Human Computer Interaction Institute in the School of Computer Science at Carnegie Mellon University. I returned to BYU in 1998. I am currently the CEO of a software startup in educational technology (SparxTeq, Inc).

4. During the course of my academic career, I authored over 70 papers in the field of computer science. The topics on which I have published papers include: User Interface Management Systems; Interaction over the Internet; Syntactic representations of user interfaces; Multi-user interaction across networks; Induction of interaction behavior from pictures; Novel interaction techniques using laser pointers; Structure of speech-based interaction and integration of speech with other forms of interaction; Interactive machine learning; Interactive robotics; and Interactive television.

5. I have extensive experience with graphical user interfaces that are driven by communications-based technologies. Out of my last 70+ published papers, 14 have involved development of custom network protocols to allow devices to interact and access information. In addition, there are 6 papers that explicitly address speech interaction and the integration of other interactive modalities with speech.

6. I currently hold 4 patents in human-computer interaction. I have authored 3 textbooks on the techniques of software design for human-computer interaction.

7. I have had extensive involvement in professional societies, such as the Association for Computing Machinery (ACM), the premier society in computing. I have served in many offices of ACM's Special Interest Group on Computer Human Interaction (SIGCHI) and currently serve as its treasurer. I have been conference chair of CHI, which is the premier conference in Computer Human Interaction. I was the founding editor of ACM's *Transactions on Computer Human Interaction*. I was a co-founder and active leader for the conference on User Interface Software and Technology (UIST) for the past 29 years. I have also served at the governor's request on the Utah Science, Technology and Research (USTAR) board, which oversees and funds state economic development efforts in technology.

8. I twice received best paper awards in intelligent user interfaces. In 2004, I was appointed to the CHI Academy for international excellence in Computer Human Interaction research. In 2007, I was recognized as one of ACM's Fellows for research in computer science and in 2012 received the CHI

Lifetime Research Award, which is the highest award in Computer Human Interaction.

9. I understand that a copy of my curriculum vitae, which includes a more detailed summary of my background, experience, and publications, is provided as Ex. 1003.

## III. SUMMARY OF OPINIONS[1]

10. The opinions contained in this Declaration are based on the documents I reviewed, my professional judgment, as well as my education, experience, and knowledge regarding graphical user interfaces.

11. In forming my opinions expressed in this Declaration, I reviewed the ’718 patent (Ex. 1001); the prosecution file history for the ’718 patent (Ex. 1004); U.S. Patent Nos. 6,742,021 (Ex. 1005) and 6,851,115 (Ex. 1007), which I understand are in the chain of applications from which the ’718 patent claims priority, and their respective prosecution histories (Exs. 1006, 1008); U.S. Provisional Application Nos. 60/124,718 (Ex. 1009), 60/124,719 (Ex. 1010), and

---

[1] My citations to non-patent publications are to the original page numbers of the publication, and my citations to U.S. Patents are to the column:line number or paragraph number of the patents or published patent applications, as applicable.

60/124,720 (Ex. 1011), to which I understand the ’718 patent claims priority; Cheyer *et al.*, “Multimodal Maps: An Agent-based Approach,” published in *Proc. of the International Conference on Cooperative Multimodal Communication (CMC/95)*, Eindhoven, The Netherlands, May 1995(“*Cheyer*”) (Ex. 1012); U.S. Patent No. 5,197,005 to Shwartz *et al.* (“*Shwartz*”) (Ex. 1013); U.S. Patent No. 5,748,974 to Johnson (“*Johnson*”) (Ex. 1014); U.S. Patent No. 6,188,985 to Thrift *et al.* (“*Thrift*”) (Ex. 1015); U.S. Patent No. 6,345,389 to Dureau (“*Dureau*”) (Ex. 1016); U.S. Patent No. 5,841,431 to Simmers (“*Simmers*”) (Ex. 1017); U.S. Patent No. 6,035,197 to Haberman *et al.* (“*Haberman*”) (Ex. 1018); Coen, M. H., “Building Brains for Rooms: Designing Distributed Software Agents,” AAAI’97/IAAI’97 *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence* (1997) (“*Coen*”) (Ex. 1020); Hodjat *et al.*, “An adaptive agent oriented software architecture,” in Lee *et al.* (eds.) PRICAI’98: Topics in Artificial Intelligence, *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, vol 1531, Springer, Berlin, Heidelberg (1998) (“*Hodjat*”) (Ex. 1021); U.S. Patent No. 5,584,024 to Shwartz (“*Shwartz-024*”) (Ex. 1022); Cheyer *et al.*, “MVIEWS: Multimodal Tools for the Video Analyst,” in *Proceedings of the 1998 International Conference on Intelligent User Interfaces*

*(IUI98)*, San Francisco, California (Jan. 1998) (Ex. 1023); Kehler *et al.*, "On Representing Salience and Reference in Multimodal Human-Computer Interaction," in *Proceedings of AAAI 1998 workshop on Representations for Multi-Modal Human-Computer Interaction*, Madison, Wisconsin (1998) (Ex. 1024); Cohen *et al.*, "An Open Agent Architecture," in *Proceedings AAAI Spring Symposium*, Stanford, California (March 1994) ("*Cohen*") (Ex. 1025); Martin *et al.*, "Information brokering in an agent architecture," in *Proceedings of the Second International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, Blackpool, Lancashire, UK (Apr. 1997) ("*Martin*") (Ex. 1026); Wyard *et al.*, "Spoken language systems – beyond prompt and response," *BT Technol. J.* vol. 14 no. 1 (Jan. 1996) ("*Wyard*") (Ex. 1027); Excerpts from Knaster, Presenting Magic Cap, A Guide to General Magic's Revolutionary Communicator Software, 1994 (Ex. 1028); Moran *et al.*, "Multimodal User Interfaces in the Open Agent Architecture," *Proc. of the 2nd International Conference on Intelligent User Interfaces (IUI '97)*, Orlando, Florida (1997) (Ex. 1029); Konstan, J. A., "State Problems in Programming Human-Controlled Devices," *IEEE Transactions on Consumer Electronics*, vol. 40, no. 4 (Nov. 1994) ("*Konstan*") (Ex. 1033); and any other materials I refer to in this Declaration in support of my opinions.

12. My opinions have also been guided by my appreciation of how a person of ordinary skill in the art would have understood the claims and the specification of the ’718 patent at the time of the alleged invention, which I have been asked to initially consider as the 1999 time frame, including and up to the March 17, 1999 date which the ’718 patent claims as priority date. My opinions reflect how one of ordinary skill in the art would have understood the ’718 patent, the prior art to the patent, and the state of the art at the time of the alleged invention.

13. As I discuss in detail below, it is my opinion that certain references disclose or suggest all the features recited in claims 1-27 of the ’718 patent.

## IV. PERSON OF ORDINARY SKILL IN THE ART

14. Based on my knowledge and experience, I understand what a person of ordinary skill in the art would have known at the time of the alleged invention. My opinions herein are, where appropriate, based on my understandings as to a person of ordinary skill in the art at that time. In my opinion, based on the materials and information I have reviewed, and based on my experience in the technical areas relevant to the ’718 patent, a person of ordinary skill in the art at the time of the alleged invention of the ’718 patent would have had at least a Bachelor’s degree in computer science, electrical engineering, or a similar

discipline, and one to two years of work experience in user interfaces for computer systems (including speech-based interfaces), networked computer systems, or a related area. More education can substitute for practical experience and *vice versa*. I apply this understanding in my analysis herein.

15. My analysis of the ’718 patent and my opinions in this declaration are from the perspective of one of ordinary skill in the art, as I have defined it above, during the relevant time frame, which I have been asked to assume is the March 17, 1999 timeframe (the filing date of Provisional Application Nos. 60/124,718, 60/124,719, and 60/124,720, from which the ’718 patent claims priority (Ex. 1001, Cover)). During this time frame, I possessed at least the qualifications of a person of ordinary skill in the art, as defined above.

## V. TECHNICAL BACKGROUND

16. In this section, I discuss the state of the art with respect to certain technologies relevant to the subject matter of the ’718 patent. In particular, during the time preceding March 1999, a person of ordinary skill in the art would have been aware of various developments in the areas of natural language processing, distributed computing, databases, multimodal input, and mobile computing, as I discuss below.

### A. Natural language processing

17. A person of ordinary skill in the art would have been aware of developments in the area of natural language processing systems prior to March 1999. For example, it was well known that users could interact with computers using natural language inputs, such as sentences in English (or another human language), e.g., as described in a paper by Wyard *et al.* from 1996 entitled "Spoken language systems – beyond prompt and response" ("*Wyard*"). (Ex. 1027, 187.) Enabling such natural language inputs was often desirable, as it allowed users to express their requirements or desires more directly and efficiently. (*Id.*)

18. In the mid-to-late 1990s, natural language input was frequently provided by way of spoken input. *Wyard* describes "a typical spoken language system architecture" as including a speech recognition component and a meaning extraction component. (*Id.*, 188, FIG. 1 (reproduced below).)

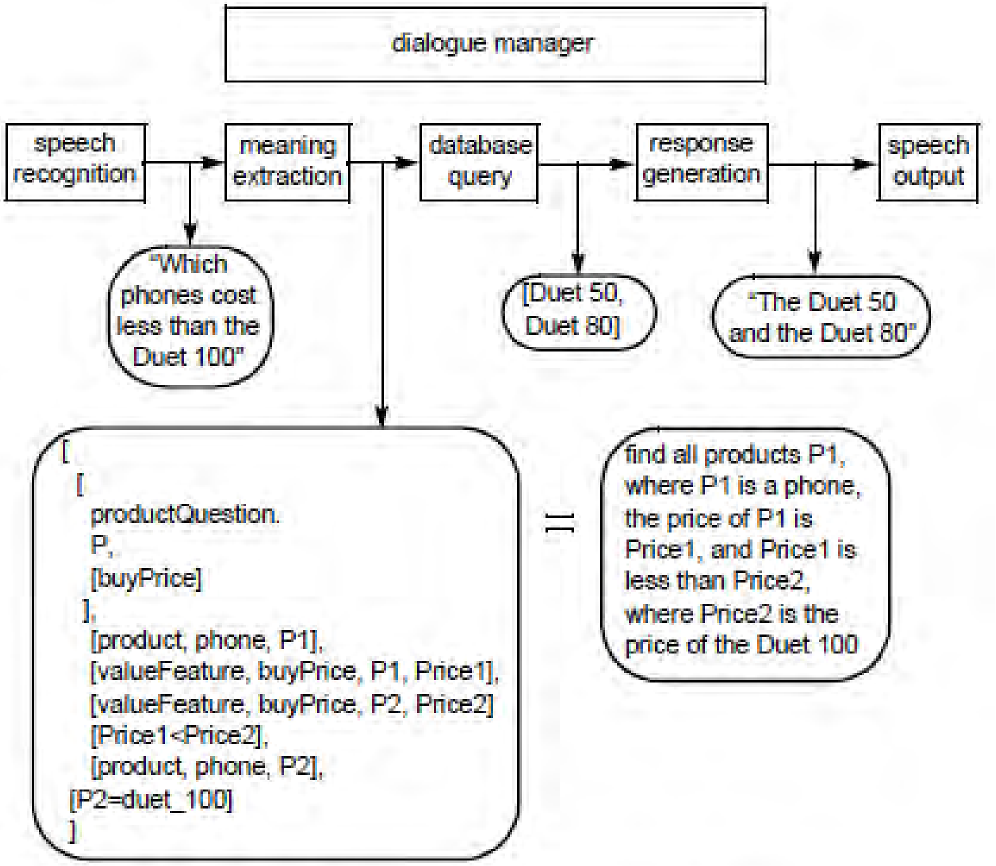


Fig 1 Example of a linear process flow in a spoken language system

(Ex. 1027, FIG. 1 (showing speech recognition and meaning extraction components that process natural language speech input.)

19. Consistent with *Wyard*'s disclosure, a person of ordinary skill would have known that the role of a speech recognition component (speech recognizer) was "to convert an input speech utterance to a string of words," and the role of a meaning extraction component was "to extract as much of the meaning as is necessary for the application from the recogniser output and encode it into a suitable meaning representation." (*Id.*, 188.)

20. A person of ordinary skill would have known that speech (voice) was one way for a user to provide natural language input, but another known way was for the user to provide language input via text, e.g., using a keyboard. It was also known prior to March 1999 that a user could use an electronic pen/stylus to provide input, e.g., by writing characters that were processed by a character recognition algorithm so that the user could enter words or sentences. For example, a paper by Moran *et al.* entitled "Multimodal User Interfaces in the Open Agent Architecture" describes input from a user via electronic pen, e.g., in conjunction with a handwriting recognizer. (Ex. 1029, 63.) The main difference between handling speech-based and text-based natural language input was that speech input had to be processed first by a speech recognizer in order to detect and identify speech utterances, whereas a speech recognizer would not have been necessary in the context of natural language text input provided via a keyboard. For handwritten text input, e.g., inputted using an electronic pen, a person of ordinary skill would have known how to implement a handwriting recognizer as discussed above.

### B. Multimodal input

21. As discussed above, a person of ordinary skill would have known before March 1999 about the existence of various input modalities, including

speech (voice), keyboard, pen/stylus, and also others such as mouse, trackball, touchpad, etc. Such a person would also have been aware of the existence and benefits of multimodal systems, which enabled a user to provide input via multiple input modalities. For example, *Wyard* describes a multimodal natural language system for providing a user with information regarding various products. (Ex. 1027, 190.) *Wyard* describes that the user can provide natural language spoken input and also click on links using a mouse and provide text as input. (*Id.*, 189 ("systems such as the BT Business Catalogue access system . . . are multimodal and require a screen and a means of inputting text and mouse clicks and outputting text and graphics."), 191 (disclosing that a user speaks "Which ones come in grey?" and later "clicks on the link next to the picture of [a particular phone that is displayed]"); *see also id.*, 190 ("a film access system, in which users will be able to select films and videos using continuous speech and button pushes on a remote control handset").) Thus, a person of ordinary skill would have been aware of "multimodal systems which aim to combine spoken language with other modalities, such as typed text and mouse clicks, in order to achieve the most user-friendly interface possible." (*Id.*, 204.)

22. As another example, a paper by *Coen* from 1997 entitled "Building Brains for Rooms: Designing Distributed Software Agents" ("*Coen*") describes an

information retrieval system with which users can interact using pointing and natural language speech input. (Ex. 1020, 975.) *Coen* discloses techniques for resolving what the user means when he/she provides the natural language spoken input "What's the weather here?" while pointing somewhere. (*Id.*) *Coen* refers to this process as "multimodal resolution." (*Id.*)

23. Thus, a person of ordinary skill in the art would have known how to implement multimodal systems (systems that enable input via multiple input modalities) in an effective, user-friendly manner prior to March 1999.

### C. Databases

24. A person of ordinary skill would have known before March 1999 that a fundamental component of an information retrieval system was a database, and that database queries could be used to retrieve information from a database. For example, *Wyard* describes a natural language based system that includes a database query as a key processing component, in order to "retrieve the information specified by the output of the meaning extraction component." (Ex. 1027, 188.) It was known to generate a database query after first processing natural language speech input with a speech recognition component and a meaning extraction component (or processing natural language text input with a meaning extraction component), as shown in the following flow diagram in *Wyard*:

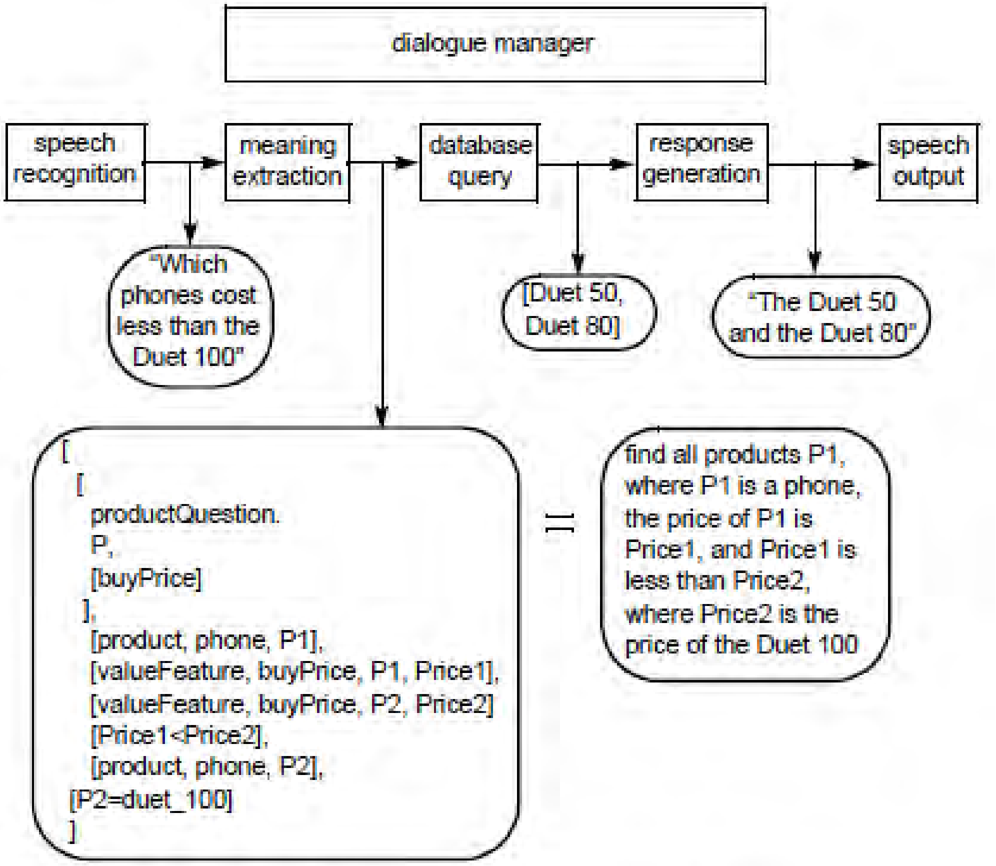


Fig 1 Example of a linear process flow in a spoken language system

(Ex. 1027, FIG. 1.)

25. *Wyard* explains the database query as follows:

> When the [dialogue manager] has prepared the query, it will be passed to the database query component. The database query component's purpose is to convert the query from the [dialogue manager] into one or more queries which can be used to find the required information from within the database. Having established the queries, the database query component then extracts the actual information from the database.

(*Id.*, 201.)

26. A person of ordinary skill would have been aware that a database query could be implemented in, e.g., structured query language (SQL), which was a well-known programming language (and one of the most prevalent and commonly used languages) for working with databases. (*Id.*, 202 ("the database querying module provides a means of separating the actual database query (in SQL, for example) from the internal representation in the [dialogue manager]"). For example, as disclosed in U.S. Patent No. 5,584,024 to Shwartz ("*Shwartz-024*") (Ex. 1022), a person of ordinary skill would have known how to use an SQL statement called a SELECT statement to retrieve a set of records from database tables. (*See* Ex. 1022, FIGS. 2C, 3A, 1:56-2:26 (disclosing examples of SQL SELECT statements).) A SELECT statement was a fundamental aspect of SQL, and similar statements were used in other database programming languages.

27. As one example, *Shwartz-024* discloses that "to produce from a database a list of customer names and phones for New York customers sorted by zip code, the following SQL statement could be used: . . . SELECT NAME, PHONE FROM CUSTOMERS WHERE STATE = `NY` ORDER BY ZIP_CODE." (*Id.*, 1:59-66.) "In this example, the SELECT command defines which fields to use, the WHERE command defines a condition by which database

records are selected, and ORDER BY keywords define how the output should be sorted." (*Id.*, 2:1-4.) "The FROM keyword defines in which tables the fields are located." (*Id.*, 2:4-5.)

28. A person of ordinary skill would have known that databases were in widespread usage across a variety of contexts long before March 1999. For example, the World Wide Web ("the Web"), which was created in the early 1990s, involved web servers that provide users with access to remote databases. The Web was in widespread usage by March 1999, and a person of ordinary skill would have known how to program computers to access information from the Web. A person of ordinary skill would have known how to implement databases available via the Web to be accessible via database queries.

### D. Distributed computing

29. A person of ordinary skill in the art would have been familiar with distributed systems for various computing contexts. Such a person would have known that various networked entities could communicate with one another and take respective actions to accomplish goals. For example, *Coen* describes "a distributed software agent system that controls the behavior of [a] laboratory's Intelligent Room." (Ex. 1020, 971 (at Abstract).) *Coen* discloses that a "system of

software agents . . . known collectively as the Scatterbrain" control various aspects of a room. (*Id.*, 971.) *Coen* explains:

> The Scatterbrain consists of approximately 20 distinct, intercommunicating software agents that run on ten different networked workstations. These agents' primary task is to link various components of the room (e.g., tracking cameras, speech recognition systems) and to connect them to internal and external stores of information (e.g., a person locator, the World Wide Web). Although an individual agent may in fact perform a good deal of computation, we will focus our interest on the ways in which agents get connected and share information rather than how they internally manipulate their own data. And while the Intelligent Room is a fascinating project in itself, we will treat it here mainly as a test-bed to learn more about how software agents can interact with other computational and real entities.

(*Id.*)

30. *Coen* discloses that "[p]eople can interact with [a system in the room called Storm] using pointing and speech." (*Id.*, 975.) For example, when the user provides a natural language spoken input "Computer, what is the weather here?" the room "displays a weather forecast for San Juan." (*Id.*) *Coen* discloses various agents, such as a SpeechIn Agent (for interfacing with speech recognition systems), Tracking Agent (for updating another agent in real-time), Weather Agent

(for obtaining forecasts and satellite maps for particular places), and Display Agent (for displaying content at a location in the room where people can see it). (*Id.*, 974.) "All the Scatterbrain agents then work together in parallel with different inputs and data being processed simultaneously in different places." (*Id.*)

31. A person of ordinary skill would have known how to implement agents in a layered, hierarchical configuration. For example, *Coen* discloses that "[layered] on top of the Scatterbrain, we created higher-level agents that rely on the Scatterbrain's underlying behaviors." (*Id.*)

32. A paper by Hodjat *et al.* from November 1998 entitled "An Adaptive Agent Oriented Software Architecture" ("*Hodjat*") describes an agent as an "autonomous individual the internals of which are not known and that conforms to a certain standard of communications and/or social laws with regard to other agents." (Ex. 1021, 33.) *Hodjat* discloses "an agent-oriented methodology, which can be universally applied to any software design." (*Id.*, 34.) A person of ordinary skill would have known, based on *Hodjat*, how to configure an agent-based architecture so that "new agents supply other agents with information about their capabilities and needs." (*Id.*, 35.) Like *Coen*, *Hodjat* describes a cooperative collection of agents that coordinate with one another, including in a hierarchical manner, to accomplish a set of requests:

> The software as a whole should be thought of as a society, striving to accomplish a set of requests. The input requests are therefore propagated, processed by agent modules that may in turn create requests to other agents. Again, it is up to the designers to break down the system, as they feel suitable. Hierarchies of agents are possible and agents can be designed to be responsible for the minutest processes in the system.

(*Id.*, 37.)

33. *Hodjat* explains that a known technique for implementing distributed systems with cooperative agents was to use the then-existing Open Agent Architecture:

> [Cheyer et al 96] use the Open Agent Architecture (OAA) . . . as a basis for their design. In this approach, based on a "federation architecture" . . . , the software is comprised of a hierarchy of facilitators and agents. The facilitators are responsible for the coordination of the agents under them so that any agent wanting to communicate with any other agent in the system must go through a hierarchy of facilitators (starting from the one directly responsible for it). Each agent, upon introduction to the system, provides the facilitator above it with information on its capabilities . . . .

(*Id.*, 40.)

34. A person of ordinary skill would have been familiar with agent-based architectures like the Open Agent Architecture and would have known how to use it to implement distributed systems in various contexts, including speech-based information retrieval. The Open Agent Architecture was described in published documents at least as early as 1994, when Cohen *et al.* described in a paper entitled "An Open Agent Architecture" ("*Cohen*") an "open agent architecture . . . served by a multimodal interface, including pen, voice, and direct manipulation" and that included "a User-interface agent that accepts spoken or typed . . . natural language queries from the user and presents responses to the queries." (Ex. 1025, 1 (at Abstract), 3.) The Open Agent Architecture was also described in several other published documents prior to March 1999. (*See, e.g.*, Ex. 1023, 57-58, Ex. 1024, 34; Ex. 1026, 472.) A person of ordinary skill would have known how to use agents to accomplish a goal in a distributed manner, e.g., based on the following disclosure in *Cheyer*:

> The architecture for the OAA, based loosely on Schwartz's FLiPSiDE system[ ], uses a hierarchical configuration where client agents connect to a 'facilitator' server. Facilitators provide content-based message routing, global data management. and process coordination for their set of connected agents. Facilitators can, in turn, be connected as clients of other facilitators. Each facilitator records the

> published functionality of their sub-agents, and when queries arrive in Interagent Communication Language form, they are responsible for breaking apart any complex queries and for distributing goals to the appropriate agents. An agent solving a goal may require supporting information and the agent architecture provides numerous means of requesting data from other agents or from the user.
>
> The Open Agent Architecture provides capability for accessing distributed knowledge sources through natural language and voice . . . .

(Ex. 1012, 7-8; *see also id.*, 9 ("In the Open Agent Architecture, agents are distributed entities that can run on different machines, and communicate together to solve a task for the user.").)

35. A person of ordinary skill would have known before March 1999 that many other agent-based architectures could also be used for implementing distributed systems. It was known at least as early as 1994 that "[a]gents are all the rage." (Ex. 1025, 1.) A person of ordinary skill would have been motivated to implement systems in a distributed manner for a variety of reasons, including increased speed, redundancy, reliability, security, and flexibility in design and implementation.

### E. Mobile computing

36. A person of ordinary skill in the art would have known prior to March 1999 that computing devices could be implemented in variety of form factors, including mobile, handheld computing devices such as personal digital assistants (PDAs) and smartphones. A person of ordinary skill would have known how to program mobile computing devices to receive information from remote data sources. For example, mobile computing devices equipped with the Magic Cap operating system were capable of retrieving information from a remote database, e.g., in the context of receiving electronic mail. (Ex. 1028, 2 ("Every Magic Cap communicator has a jack where you can plug in a telephone line. This is how you'll use your communicator to send and receive electronic mail . . . .").)

37. A person of ordinary skill would also have known how to implement agent-based distributed software systems on a mobile computing device. For example, *Cohen* described in 1994 three types of agent-based software systems implemented in the Apple Newton, which was a PDA. (Ex. 1025, 1 ("Each of [three general conceptions of agent-based software systems] can be found to some extent in present-day software products, for example, in . . . Apple Computer's Newton . . . .")

## VI. OVERVIEW OF THE '718 PATENT

38. The ’718 patent “relates generally to the navigation of electronic data by means of spoken natural language requests, and to feedback mechanisms and methods for resolving the errors and ambiguities that may be associated with such requests.” (Ex. 1001, 1:22-26.) Figure 4, reproduced below, depicts an exemplary process in accordance with one embodiment of the ’718 patent.

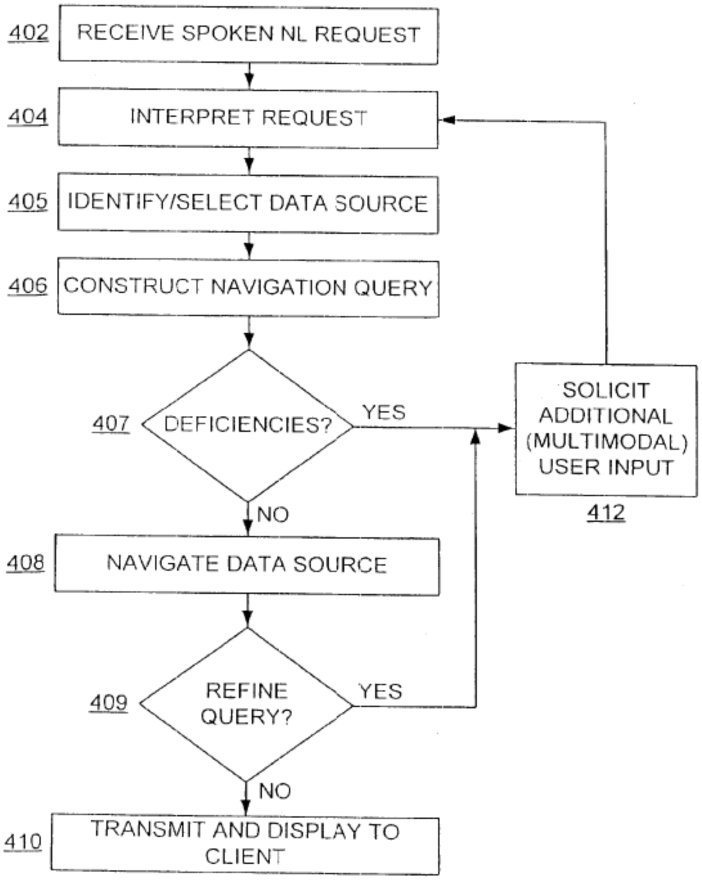


Fig. 4

(Ex. 1001, FIG. 4.)

39. The process depicted in Figure 4 begins at step 402, where "the user's spoken request for information is initially received in the form of raw (acoustic) voice data." (*Id.*, 7:19-22, FIG. 4.) "At step 404 the voice data received from the user is interpreted in order to understand the user's request for information." (*Id.*, 7:22-24, FIG. 4.) "In step 405 request processing logic 300 identifies and selects an appropriate online data source where the desired information . . . can be found." (*Id.*, 8:51-54, FIGS. 1A-1B (showing data sources 110), FIG. 4.)

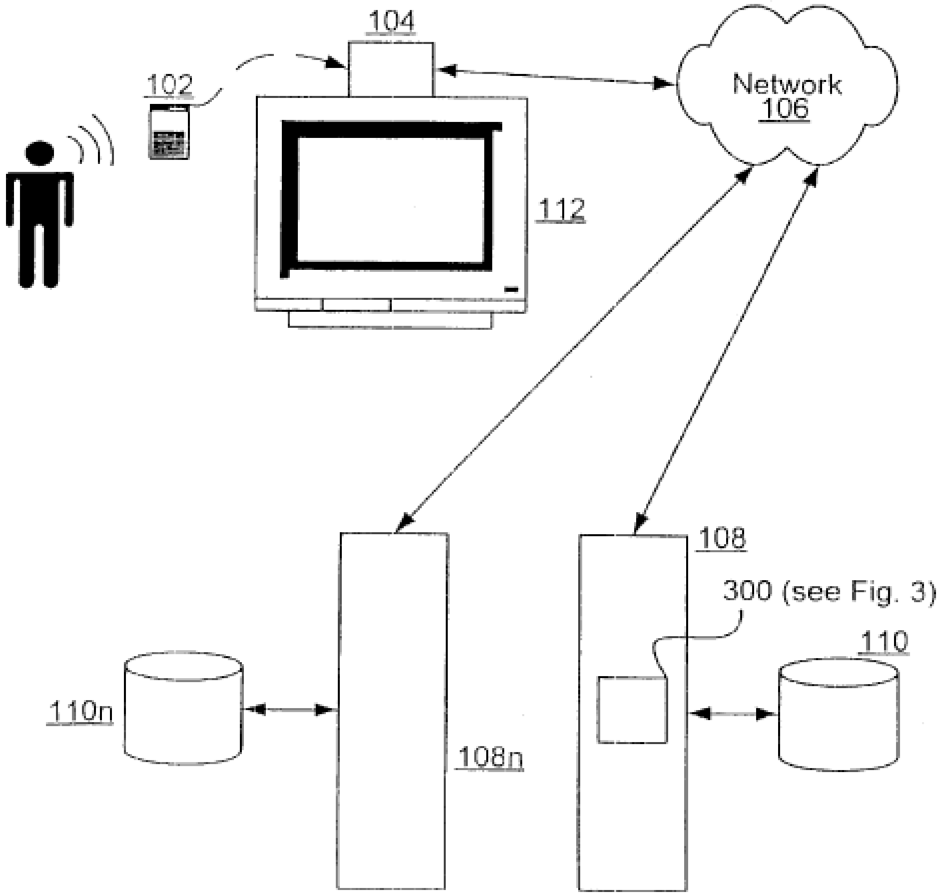


Fig. 1a

“Step 406 attempts to construct a navigation query, reflecting the interpretation of step 404.” (*Id.*, 8:62-63, FIG. 4.) At step 407, “deficiencies may be identified during the process of query construction” (*id.*, 10:51-54, FIG. 4.), in which scenario “additional input [is solicited] from the user . . . via user interface modalities in addition to spoken natural language (‘multi-modality’)” to handle “errors and deficiencies in user input.” (*Id.*, 11:16-21, FIG. 4.)

40. "Step 408 navigates the data source using that query and retrieves the desired . . . information" from an electronic data source." (*Id.*, 12:51-53, FIG. 4; *see also id.*, 9:67-10:2 ("the query thus constructed . . . is used to navigate the online data source in step 408").) "Step 409 detects that additional user input is needed to further refine the query in order to select a particular film for viewing," and this is another scenario in which step 412 (soliciting additional input) will be performed. (*Id.*, 11:62-64, FIG. 4.) The retrieved information is "transmitted in step 410 from network server 108 to client display device 112 via communications network 106." (*Id.*, 12:16-19, FIGS. 1A, 1B, 4.)

41. The '718 patent discloses using the then-existing Open Agent Architecture (OAA) in various embodiments. (*Id.*, 3:46-48 ("FIG. 6 illustrates an embodiment of the present invention utilizing a community of distributed, collaborating electronic agents."), 13:16-19, 14:27-29, FIG. 6 (reproduced below).) The Open Agent Architecture includes multiple "autonomous entities, or agents" and a facilitator agent. (Ex. 1007[2], 4:20-21; Ex. 1001, FIG. 6 (reproduced below).)

---

[2] Application No. 09/225,198, which issued as U.S. Patent No. 6,851,115 (Ex. 1007), is incorporated by reference into the '718 patent. (Ex. 1001, 1:5-18, 13:19-22.)

The agents forward service requests to the facilitator, which interprets such requests, organizing a set of goals which are then delegated to appropriate agents for task completion. (Ex. 1007, 6:10-13; Ex. 1001, 13:34-51.)

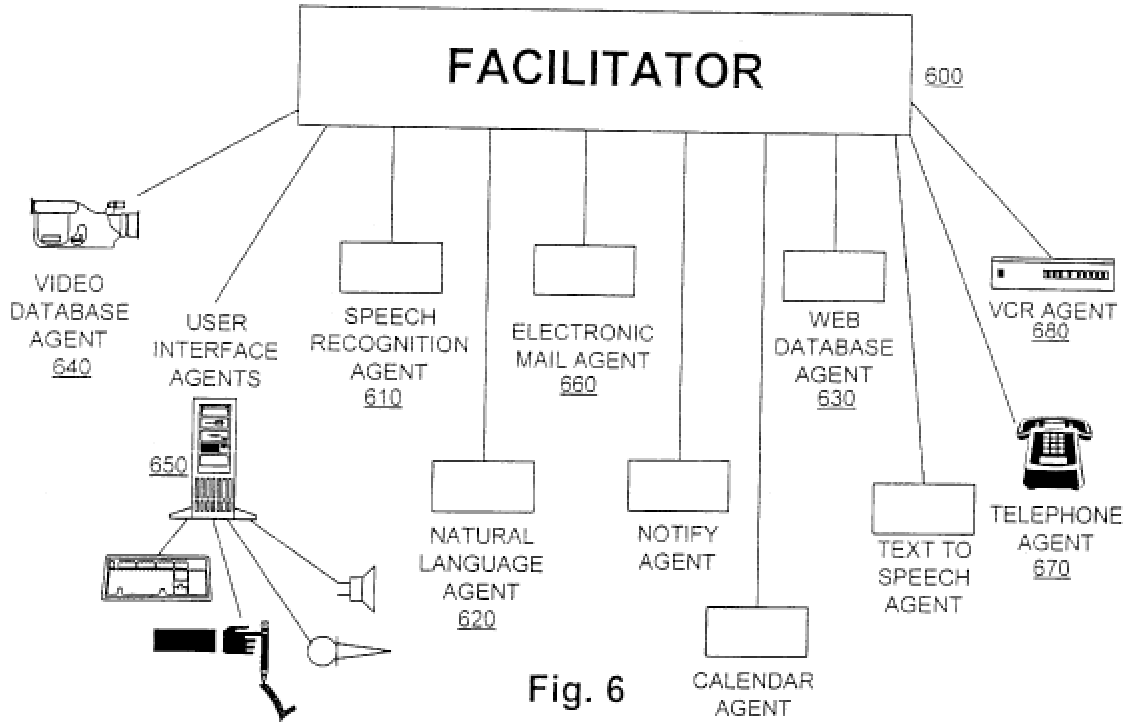


Fig. 6

(Ex. 1001, FIG. 6.)

42. The ’718 patent discloses that “an agent registers with its parent facilitator a specification of the capabilities and services it can provide,” and “[w]hen a facilitator determines that the registered capabilities of one of its client agents will help satisfy a current goal or sub-goal thereof, the facilitator delegates that sub-goal to the client agent . . . .” (Ex. 1001, 13:36-45.)

## VII. CLAIM CONSTRUCTION

43. Except for claim terms that I have identified explicitly in this section, I have given all the claim terms of the challenged claims their ordinary and customary meaning, as would be understood by a person of ordinary skill in the art, at the time of the alleged invention, which I understand is the early part of 1999 (including March 17, 1999, the claimed priority date of the ’718 patent) having taken into consideration the language of the claims, the specification, the drawings, and the prosecution history of record.

### A. “navigation query”

44. I have been asked to assume that the claim term “navigation query” recited in claims 1, 4, 10, 13, 19, and 22 is to be construed as “an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information.” I agree that this construction aligns with the disclosure in the specification of the ’718 patent that “[a] ‘navigation query’ means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information.” (Ex. 1001, 8:65-9:1.) I have applied this understanding in my analysis.

### B. “code segment [that] . . . ”

45. I have been asked to assume that the construction of the phrases in claims 10 and 13 of the form "code segment [that] [performs a function]" includes software running on a microprocessor configured to perform the functions recited in each of those phrases or equivalents thereof. In particular, claim 10 recites "code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user" (the recited function is "receives a spoken request for desired information from the user utilizing the mobile information appliance of the user"), "code segment that renders an interpretation of the spoken request" (the recited function is "that renders an interpretation of the spoken request"), "code segment that constructs a navigation query based upon the interpretation" (the recited function is "constructs a navigation query based upon the interpretation"), "code segment that utilizes the navigation query to select a portion of the electronic data source" (the recited function is "utilizes the navigation query to select a portion of the electronic data source"), and "code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user" (the recited function is "transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user"), and claim 13 recites "code segment that solicits additional input from the user,

including user interaction in a modality different than the original request" (the recited function is "solicits additional input from the user, including user interaction in a modality different than the original request"), "code segment that refines the navigation query, based upon the additional input" (the recited function is "refines the navigation query, based upon the additional input"), and "code segment that uses the refined navigation query to select a portion of the electronic data source" (the recited function is "uses the refined navigation query to select a portion of the electronic data source"). I have applied this understanding in my analysis.

### C. **"... logic[,] operable to ... "**

46. I have been asked to assume that the construction of the phrases in claims 19 and 22 of the form ". . . logic[,] operable to [perform a function]" includes software running on a microprocessor configured to perform the functions recited in each of those phrases or equivalents thereof. In particular, claim 19 recites "spoken language processing logic, operable to render an interpretation of the spoken request" (the recited function is "render an interpretation of the spoken request"), "query construction logic, operable to construct a navigation query based upon the interpretation" (the recited function is "construct a navigation query based upon the interpretation"), and "navigation logic, operable to select a portion

of the electronic data source using the navigation query" (the recited function is "select a portion of the electronic data source using the navigation query"), and claim 22 recites "user interaction logic operable to solicit additional input from the user, including user interaction in a modality different than the original request" (the recited function is "solicit additional input from the user, including user interaction in a modality different than the original request") and "query refining logic operable to refine the navigation query based upon the additional input" (the recited function is "refine the navigation query based upon the additional input"). I have applied this understanding in my analysis.

## VIII. OVERVIEW OF THE PRIOR ART

### A. *Cheyer*

47. *Cheyer*, whose authors are two of the named inventors of the '718 patent, describes "how multiple input modalities may be combined to produce more natural user interfaces." (Ex. 1012, 1 (at Abstract).) *Cheyer* discloses a "map-based application for a travel planning domain" that is "distinguished by a synergistic combination of handwriting, gesture and speech modalities; access to existing data sources including the World Wide Web; and a mobile handheld interface." (*Id.*) *Cheyer*'s multimodal application uses the then-existing Open

Agent Architecture to implement "a distributed network of heterogeneous software agents" for distributed processing regarding various tasks. (*Id.*)

48. *Cheyer* discloses various examples of receiving a spoken natural language (e.g., English) request for desired information from a user, such as: "'Display the French restaurants within 1 mile of this hotel,'" "'Show me all available information about Alcatraz,'" and "'What is the distance from the post office to the hotel?'" (Ex. 1012, 4 ("spoken natural language"), 5.) *Cheyer* discloses that the user's computing device, which may be a PC or a handheld PDA, receives the spoken request, e.g., using a microphone for voice input. (Ex. 1012, 4, 6.)

49. The spoken English request is processed by a speech recognition (SR) agent and a natural language (NL) parser agent to recognize a speech string in the user's speech input and translate the recognized request into a format called Interagent Communication Language that software agents can handle. (Ex. 1012, 7, 9-11.) The SR and NL agents are among several agents (shown below in Figure 3 of *Cheyer*) that are implemented using the Open Agent Architecture to perform various tasks to service the user's request. (Ex. 1012, 7-12.)

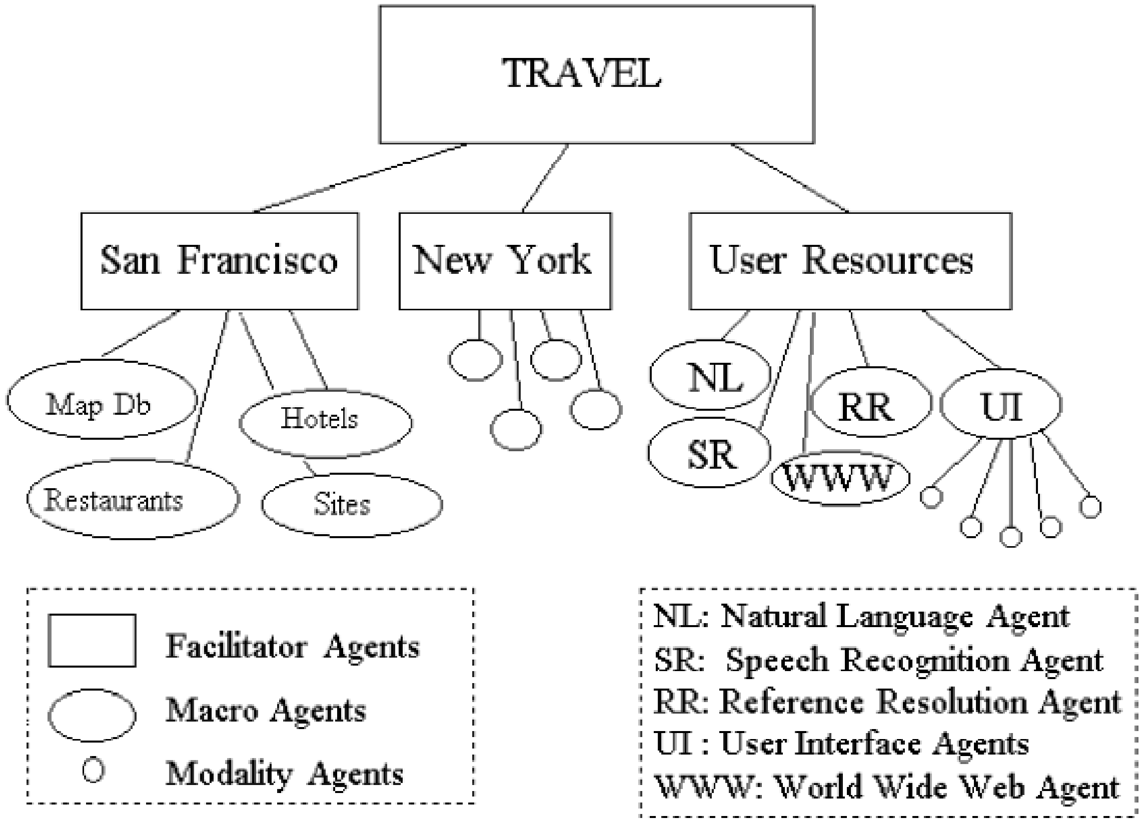


Figure 3: Agent Architecture for Map Application

(Ex. 1012, FIG. 3.)

50. *Cheyer* discloses that "[t]he architecture for the OAA . . . uses a hierarchical configuration where client agents connect to a 'facilitator' server," also referred to as a "facilitator agent." (Ex. 1012, 7, 9.) *Cheyer* discloses that the facilitator agent "records the published functionality of [its] sub-agents." (Ex. 1012, 8.)

51. *Cheyer* discloses that in one example, the user may issue a request via a "synergistic combination of pen and voice, by speaking 'What is the distance from here to this hotel?' while simultaneously indicating the specified locations by pointing or circling." (Ex. 1012, 5-6.) In another example, the "user speaks: 'How far is the restaurant from this hotel?'" but has not yet indicated what is "this hotel," so the user's request is "ambiguous or underspecified." (*Id.*, 6 ("ambiguous or underspecified"), 11-12.) *Cheyer* discloses that "the system will wait several seconds and then issue a prompt requesting additional information." (Ex. 1012, 6.) For example, a "reference resolution agent (RR)" asks for resolution of an unclear reference such as "this hotel." (*Id.*, 12.) "The interface agent . . . waits for the user to make a gesture indicating '[this] hotel', issuing prompts if necessary." (*Id.*) *Cheyer* discloses that after unclear references have been resolved, a domain agent "sends database requests" asking for information from a database relevant to servicing the user's request. (*Id.*) The domain agent then "requests the user interface to produce output" responsive to the user's spoken request. (*Id.*)

### B. *Shwartz*

52. *Shwartz* relates to a "database retrieval system having a natural language interface." (Ex. 1013, Title; *see also id.*, 1:9-11.) *Shwartz*'s system includes a query system that "allows users with little or no computer experience to

enter a conversational English (or other natural language) query" and a "natural language interface [that] interprets the query and reduces it into an internal meaning representation used by the system," e.g., using a natural language parser. (*Id.*, 5:60-62, 6:3-7, 7:38-41.) *Shwartz* discloses that "data responsive to the query is located using a database expert system that enables retrieval of the data from proper tables and columns in the database." (*Id.*, 6:11-14.) *Shwartz* further discloses a "navigator and query language generator 38 [that] is used to define optimal navigation paths through the database tables and columns to respond to the query, and to generate a meta-query language ('MQL')," and a reporter and database access system 40 that uses the meta-query language "to generate the code (e.g., structured query language ('SQL') code) to actually retrieve the information from the application database." (*Id.*, 9:28-35.)

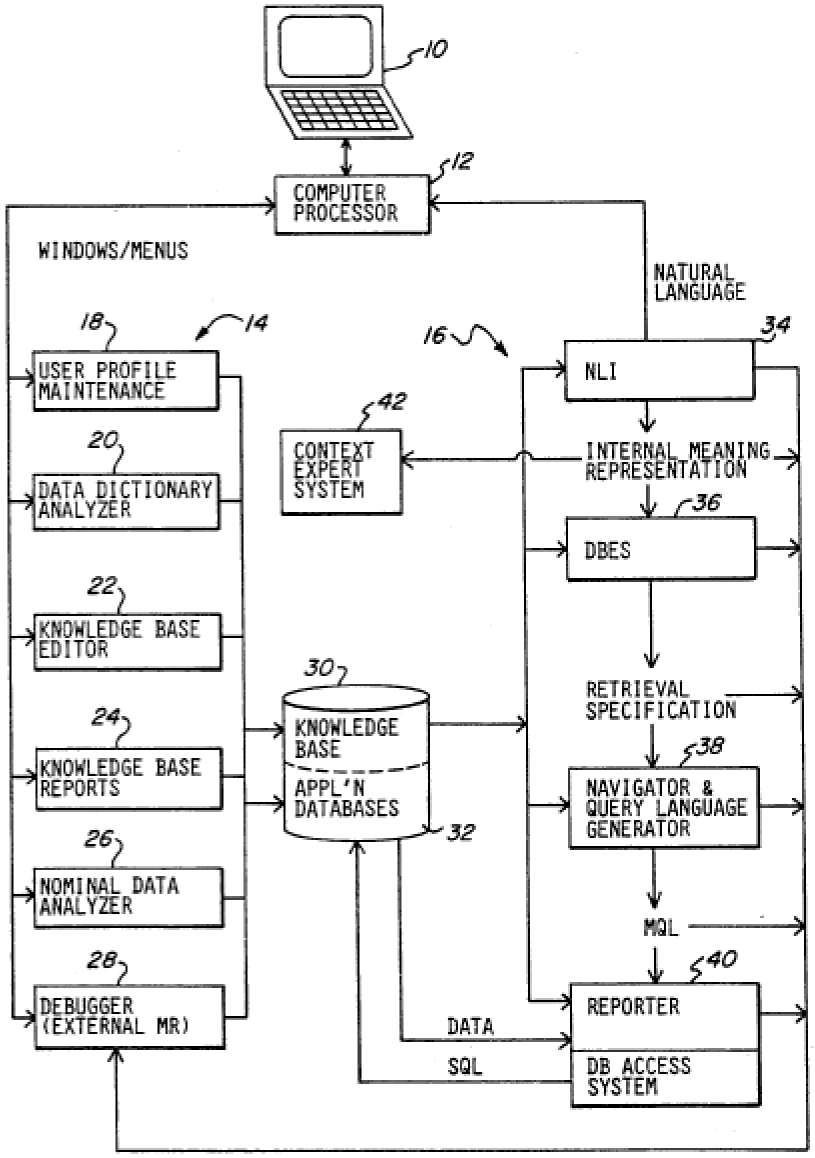


*FIG. 1*

(*Id.*, FIG. 1 (showing natural language interface 34, database expert system 36, navigator and query language generator 38, and reporter and database access system 40).)

53. Thus, *Shwartz*, like *Cheyer*, discloses a natural language based information retrieval system, and further discloses generating a query (e.g., an SQL query) for retrieving information from a database.

### C. *Johnson*

54. *Johnson* relates to a "multimodal natural language interface [that] interprets user requests combining natural language input from the user with information selected from a current application." (Ex. 1014, Abstract.) *Johnson* discloses a system that accepts "user input [that] may be spoken, typed, handwritten, mouse controlled cursor, touch, or any other modality." (*Id.*, 3:44-46.) *Johnson* discloses that speech input is processed by a speech recognizer 41, and "output of the speech recognizer 41 and the non-speech input received by the screen manager 42 are sent to a dispatcher 44 which combines the inputs and directs the combined input to first of all a natural language processor 45." (*Id.*, 3:63-67.) The combined multimodal input is parsed at a parser/semantic interpreter 46. (*Id.*, 3:67-4:2.)

55. *Johnson* discloses that in the example of a database query for Joe Smith's telephone number, there could be two Joe Smiths in the database, so that "there is an ambiguity that must be clarified before a final response can be generated." (*Id.*, 5:7-18; *see also id.*, FIG. 4 (reproduced below).) If there is an

ambiguity, *Johnson*'s system asks the user to select one of the possibilities or indicate whether to look elsewhere. (*Id.*, FIG. 4.)

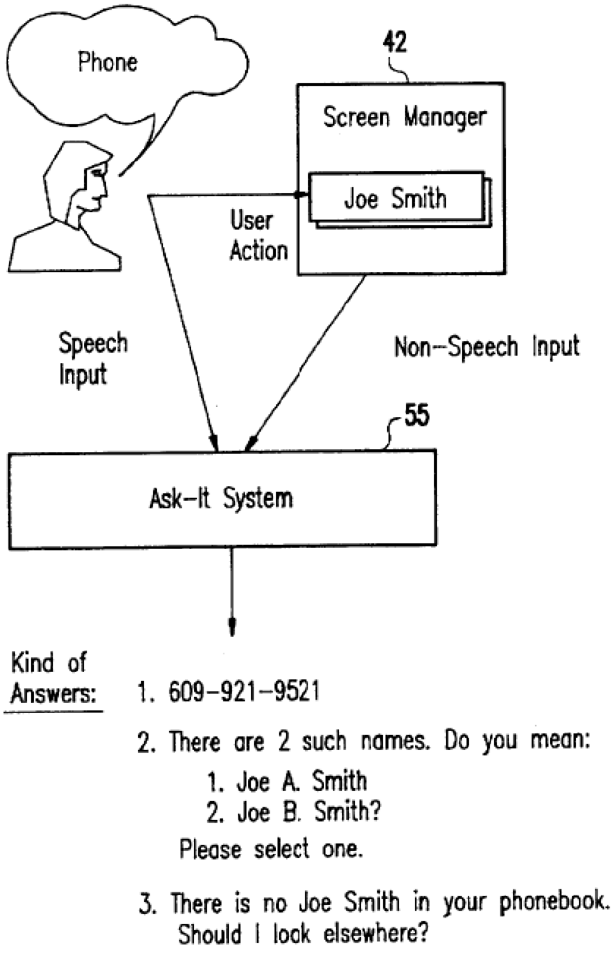


FIG.4

(Ex. 1014, FIG. 4.)

56. Thus, *Johnson*, like *Cheyer* and *Shwartz*, discloses a natural language based information retrieval system, and like *Cheyer*'s system, *Johnson*'s system

includes capabilities for receiving multimodal input and for clarifying an ambiguous user request.

### D. *Thrift*

57. *Thrift* "relates generally to voice recognition devices, and more particularly to a wireless voice-controlled device that permits a user to browse a hypermedia network, such as the World Wide Web, with voice commands." (Ex. 1015, 1:10-14; *see also id.*, 2:37-39 ("The invention described herein is directed to a wireless voice-activated device for controlling a processor-based host system.").) *Thrift* explains that "[i]n the example of this description, the host system is a computer connected to the World-Wide Web and the device is used for voice-controlled web browsing[;] [h]owever, the same concepts can be applied to a voice-controlled device for controlling any processor-based system that provides display or audio information, for example, a television." (*Id.*, 2:40-46.)

58. *Thrift* discloses that "[a]n example of voice control interpretation other than for Web browsing is for commands to a television, where host system 11 is a processor-based television system." (*Id.*, 3:57-59.) "For example, the vocal command, 'What's on TV tonight?', would result in a display of the television schedule." (*Id.*, 3:59-60.) *Thrift* also describes that "[a]nother example of voice control interpretation other than for Web browsing is for commands for computer-

based household control," in which context "[t]he vocal command, 'Show me the sprinkler schedule' would result in an appropriate display." (*Id.*, 3:61-65.) Thus, *Thrift* discloses various examples of providing information to a user based on voice input and is therefore in the same field as *Cheyer*.

59. Figure 1 of *Thrift* "illustrates one embodiment of a wireless voice-activated control unit 10 in accordance with the invention." (*Id.*, 2:54-55, FIG. 1 (reproduced below).)

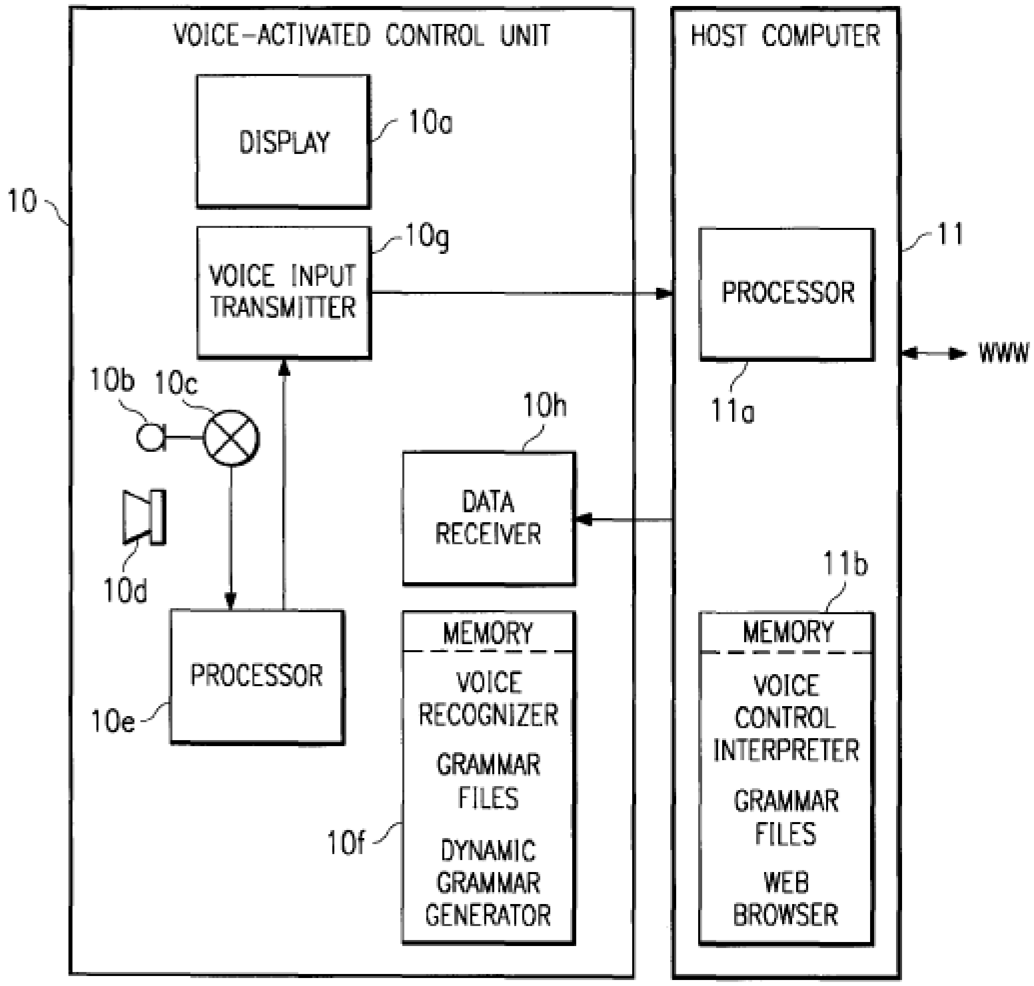


*FIG. 1*

(Ex. 1015, FIG. 1 (showing a voice-activated control unit 10 and a host computer 11).)

### E. *Dureau*

60. *Dureau* "relates generally to interactive television systems" (Ex. 1016, 1:8-12) and discloses voice input to a set-top box coupled to a television. (*Id.*, Abstract ("[A] microphone is coupled to a set-top box. The microphone allows the

user to input voice information which is digitized and conveyed to the server for conversion into textual information."), 10:56-11:1 ("[T]he user can enter his information by voice. The user can use a microphone or a telephone handset to provide voice data to the system. The microphone may [be] a special-purpose microphone for use with the interactive television system or it may be a telephone handset. A special-purpose microphone may be connected to the set-top box, or it may be built into a remote control for the system. A telephone handset may be connected to the set-top box, or it may be connected directly to the return path (i.e., telephone line.) The voice data is transmitted to the server, which uses voice recognition software to convert the voice data into textual data. The textual data is returned to the set-top box, where it can be displayed to the user."), FIG. 1 (reproduced below).)

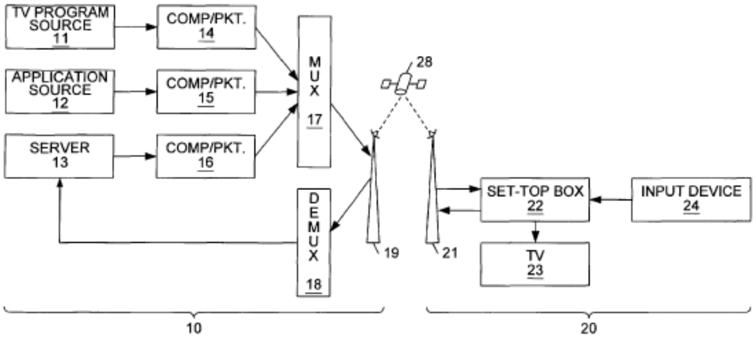


FIG. 1

(Ex. 1016, FIG. 1 (showing set-top box 22 connected to television 23).)

### F. *Simmers*

61. *Simmers* "relates to graphical displays connected to information devices." (Ex. 1017, 1:9-10.) *Simmers* discloses a "dual-function information devices such as a cellular phone with PDA." (*Id.*, 1:47-48; *see also id.*, 1:12-15 ("'smart' cellular phones, which function both for telecommunications and for storing and retrieving information (e.g., a Personal Digital Assistant (information device))".)

## IX. THE PRIOR ART DISCLOSES OR SUGGESTS ALL OF THE FEATURES OF CLAIMS 1-27 OF THE '718 PATENT

### A. *Cheyer*, *Shwartz*, and *Thrift* Disclose or Suggest the Features of Claims 1-4, 6, 8-10, 12, 13, 15, 17-19, 21, 22, 24, 26, 27

62. I reviewed *Cheyer*, *Shwartz*, and, *Thrift*, and in my opinion, *Cheyer* and *Shwartz* disclose or suggest all of the features of claims 1-4, 6, 8-10, 12, 13, 15, 17-19, 21, 22, 24, 26, 27of the '718 patent. Below, I address each of these claims and their respective limitations.

### 1. Claim 1

#### i) "A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:"

63. I have been asked to assume that the preamble of claim 1 is limiting. Under that assumption, it is my opinion that *Cheyer* discloses the limitations in the preamble of claim 1.

64. For instance, *Cheyer* discloses a method for processing input provided by a user via "spoken natural language" (Ex. 1012, 4) ("speech-based") to enable the user "to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web" (*id.*) ("navigation of an electronic data source"). (*See also id.*, 11-12 (providing an example where a user's speech-based query is processed to provide the user with requested information).)

65. More specifically as to "*speech-based*," *Cheyer* discloses an "application [that] is distinguished by a synergistic combination of handwriting, gesture and *speech* modalities." (Ex. 1012, 1 (emphasis added).) In particular, *Cheyer* provides the user with the ability to enter natural language input via a variety of modalities, including speech-based, and explains benefits associated

with such a speech-based method. (*Id.*, 2 ("Natural language content can be entered through different input modalities, including . . . speech."), 3 ("Spoken language is the modality used first and foremost in human-human interactive problem solving . . . . Speech is an extremely fast medium, several times faster than typing or handwriting. In addition, speech input contains content that is not present in other forms of natural language input, such as prosody, tone and characteristics of the speaker (age, sex, accent).").)

66. To process the user's speech input, *Cheyer* makes use of "[e]xisting . . . natural language and speech recognition systems." (Ex. 1012, 4; *see also id.*, 7 ("Several natural language systems have been integrated into the OAA which convert English into the Interagent Communication Language [and in] addition, a speech recognition agent has been developed to provide transparent access to the Corona speech recognition system."), 9 ("Speech Recognition (SR) Agent: . . . is also responsible for supervising a child micro agent whose task is to control the speech data stream.").) *Cheyer* provides various examples of spoken input requests by a user. (*See, e.g.*, *id.*, 5 ("The user may ask the map to perform various actions. For example, distance calculation: e.g. 'How far is the hotel from Fisherman's Wharf?' object location: e.g. 'Where is the nearest post office?' filtering: e.g. 'Display the French restaurants within 1 mile of this hotel.'

information retrieval: e.g. ‘Show me all available information about Alcatraz.’ . . . During input, requests can be entered using . . . voice. . . . [I]n order to calculate the distance between two points on the map, a command may be issued using: . . . voice, by speaking ‘What is the distance from the post office to the hotel?’”), 6 (“synergistic combination of pen and voice, by speaking ‘What is the distance from here to this hotel?’. . . . vocalization of the request to calculate the distance . . . . a microphone or a telephone for voice input . . . . The result is a mobile system that provides a synergistic pen/voice interface to remote databases. . . . Solutions to verbal commands are displayed in three to five seconds after the end of the speech has been detected; partial feedback indicating the current status of the speech recognition is provided earlier.”), 11 (“A user speaks: ‘How far is the restaurant from this hotel?’”).)

67. More specifically as to “*navigation of an electronic data source*,” *Cheyer* discloses navigation of data sources such as remote databases on the World Wide Web. (Ex. 1012, 1 (at Abstract, disclosing “access to existing *data sources* including the World Wide Web”) (emphasis added), 6 (“The interface is connected either by modem or ethernet to a server machine which will manage database access. . . . The result is a mobile system that provides a synergistic pen/voice interface to *remote databases*.”) (emphasis added), 7 (“Through the use of agents,

the OAA provides distributed access to commercial applications, such as mail systems, calendar programs, databases, etc.”), 10 (describing types of databases that are used), 12 (“[T]he domain agent (RR) sends database requests asking for the coordinates of the items in question. . . . The resulting application has met our initial requirements: a mobile, synergistic pen/voice interface providing good natural language *access to heterogeneous distributed knowledge sources*”) (emphasis added).)

68. *Cheyer* discloses that the remote database is located at one or more network servers located remotely from a user. For example, *Cheyer* discloses “access to existing data sources including the World Wide Web” (Ex. 1012, Abstract), and explains that its system enables “a mobile system that provides a synergistic pen/voice interface to remote databases” (*id.*, 6). A person of ordinary skill would have understood that the way a user’s device retrieved information from the World Wide Web was by contacting a remote server (e.g., web server) that could transmit the information to the user’s device. Indeed, the existence of servers on a network that enabled a user to access data remotely was one of the fundamental principles of the World Wide Web.

69. A person of ordinary skill would have understood that *Cheyer* necessarily discloses that a data link is established between the user’s mobile

device ("mobile information appliance of the user") and the remote server ("one or more network servers"). A "handheld PDA" (Ex. 1012, 4, 6) with a "mobile handheld interface" (*id.*, Abstract) as disclosed by *Cheyer* is a "mobile information appliance of the user" as recited in the preamble. *Cheyer* discloses that the "mobile system [] provides [an] interface to remote databases," and thus discloses that the user's mobile device communicates with the remote databases. (*Id.*, 6; *see also id.*, Abstract ("access to existing data sources including the World Wide Web; and a mobile handheld interface"), 4 ("Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web"), 7 ("access to various heterogeneous data and knowledge sources"), 12 ("mobile … interface providing … access to heterogeneous distributed knowledge sources").) Such communication reflects a data link between the user's mobile device and the remote server. (*Id.*, 6.)

70. (*See also* below at Sections IX.A.1.ii-vi regarding the remaining limitations of this claim.)

### ii) [1.a] "(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;"

71. *Cheyer* in combination with *Thrift* discloses this limitation. For instance, *Cheyer* discloses various examples of receiving a spoken request for desired information from a user:

- "'How far is the hotel from Fisherman's Wharf?'" (Ex. 1012, 5);
- "'Where is the nearest post office?'" (*id.*);
- "'Display the French restaurants within 1 mile of this hotel'" (*id.*);
- "'Show me all available information about Alcatraz'" (*id.*);
- "'What is the distance from the post office to the hotel?'" (*id.*);
- "'What is the distance from here to this hotel?'" (*id.*, 6); and
- "A user speaks: 'How far is the restaurant from this hotel?'" (*id.*, 11)

72. In each of these examples, the user is requesting desired information via a spoken request, because *Cheyer* discloses that input may be provided via voice.

73. *Cheyer* discloses that the user's computing device receives the spoken request. (Ex. 1012, 4 ("In order to provide the most natural user interface possible, the system permits the user to [provide] spoken natural language . . . . The user interface must be light and fast enough to run on a handheld PDA"), 6 ("The user interface runs on pen-equipped PC's or a Dauphin handheld PDA . . . using either a microphone or a telephone for voice input.").)

74. *Cheyer* also discloses that a micro agent associated with a speech recognition agent receives the spoken request after it is received by the user's computing device. (*See* Ex. 1012, 9 ("[The speech recognition] macro agent is . . . responsible for supervising a child micro agent whose task is to control the speech data stream. The SR agent can provide feedback to an interface agent about the current status and progress of the micro agent (e.g. 'listening', 'end of speech detected', etc.)"), 11 (disclosing that "[a] user speaks: 'How far is the restaurant from this hotel?'" and "[t]he speech recognition agent monitors the status and results from its micro agent . . . ."); *see also id.*, 7 ("a speech recognition agent has been developed to provide transparent access to the Corona speech recognition system"), 9 ("Micro Agents: are responsible for handling a single input . . . data stream . . . .").)

75. *Cheyer* discloses that the device that receives voice input from the user is a portable device. (Ex. 1012, Abstract ("mobile handheld interface"), 4 ("handheld PDA"), 6 ("mobile system"), 12 ("mobile . . . interface").) *Cheyer* further discloses that the user's mobile device communicates with a remote server to cause the remote server to retrieve information responsive to a user's query (e.g., "Show me all available information about Alcatraz") and send such retrieved information to the user's device, e.g., so that the user can see all available

information about Alcatraz. (*Id.*, 5; *see also id.*, 4 ("Through the multimodal interface, a user must be able to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web"), 6 ("mobile system that provides [an] interface to remote databases"), Abstract ("access to existing data sources including the World Wide Web; and a mobile handheld interface"); *see* above at Section IX.A.1.i (citations and analysis regarding data link and network server located remotely from a user); *see* below at Sections IX.A.1.v-vi.) Because the user's mobile device in *Cheyer*'s disclosure remotely causes a server to take prescribed actions (e.g., retrieve requested information and send it to the mobile device), the mobile device is a remote control device. *Cheyer* further discloses that the user's mobile device can be a PDA (Ex. 1012, 4, 6), and thus discloses a *portable* remote control device.

76. While *Cheyer* does not expressly disclose that "said mobile information appliance comprises a . . . remote control device or a set-top box *for a television*" as recited in limitation [1.a], a person of ordinary skill would have been motivated in view of *Thrift* to modify *Cheyer*'s process to include such features.

77. *Thrift* "relates generally to voice recognition devices" and discloses examples of voice-activated devices for controlling a processor-based host system. (Ex. 1015, 1:9-10; *see also id.*, Abstract ("hand-held wireless voice-activated

device (10) for controlling a host system (11), such as a computer connected to the World Wide Web."), 2:42 ("the device is used for voice-controlled web browsing"), 2:43-46 ("the same concepts can be applied to a voice-controlled device for controlling any processor-based system that provides display or audio information, for example, a television").) Thus, *Thrift* is in the same technical field as *Cheyer* (e.g., voice interface for retrieving information desired by a user). (Ex. 1012, Abstract.)

78. A person of ordinary skill implementing *Cheyer*'s process and system would have had reason to consider the teachings of *Thrift* for enhancing the feature set and functionality of *Cheyer*'s process and system. *Thrift* describes a system that "makes information on the Web more accessible and useful" and explains that "[s]peech control brings added flexibility and power to the Web interface and makes access to information more natural," and a person of ordinary skill would have recognized those attributes as being pertinent to *Cheyer*'s process, which similarly involves a voice interface for retrieving information from the Web. (Ex. 1015, 2:15-18; *see also* above at Section IX.A.1.i (citations and analysis regarding *Cheyer*'s voice interface for retrieving information from the Web).)

79. Additionally, a person of ordinary skill would have found *Thrift*'s disclosure of a system that interprets a user's command such as "What's on TV

tonight" or "Give me the weather" to be similar to *Cheyer*'s disclosure of a system that provides information to the user based on spoken commands. (Ex. 1015, 3:60, 4:58; Ex. 1012, Abstract, 4-6, 9-11; *see also* Ex. 1015, 4:25-26 ("Another speakable command is, 'Show me my speakable command list'."), 4:41-42 ("Another speakable command is 'Show me my speakable hotlist'"), 4:57-58 ("'How does the weather look today?'").)

80. Having looked to *Thrift*, a person of ordinary skill would have seen that *Thrift* discloses a wireless "voice-activated remote control device." (Ex. 1015, 2:39-40; *see also id.*, 1:66-67 ("wireless voice-activated control unit for controlling a processor-based host system"), 2:37-39 ("wireless voice-activated device for controlling a processor-based host system").) *Thrift* further discloses a remote control device in the context of controlling a television. (*Id.*, 2:43-46 ("the same concepts can be applied to a voice-controlled device for controlling any processor-based system that provides display or audio information, for example, *a television*") (emphasis added).)

81. A person of ordinary skill would have been motivated in light of the teachings of *Thrift* to configure *Cheyer*'s process and system so that the handheld device that receives input from the user ("said mobile information appliance") comprises a portable remote control device for a television. For example, a person

of ordinary skill would have recognized that just like *Cheyer*'s handheld PDA which receives speech input, *Thrift*'s voice-activated control unit 10 is wireless and includes a processor, memory, display, and a microphone to receive voice input. (Ex. 1015, 2:37 ("wireless"), 3:10-11 ("control unit 10 has a processor 10e"), 3:11-12 ("Memory 10f stores voice recognition programming to be executed by processor 10e."), 2:59-62 ("Control unit 10 has a display 10a and a microphone 10b. Display 10a is designed for compactness and portability, and could be an LCD. Microphone 10b receives voice input from a user."), Abstract ("The device (10) has a display (10a), a microphone (10b), and a wireless transmitter (10g) and receiver (10h)."), FIG. 1 (reproduced below).)

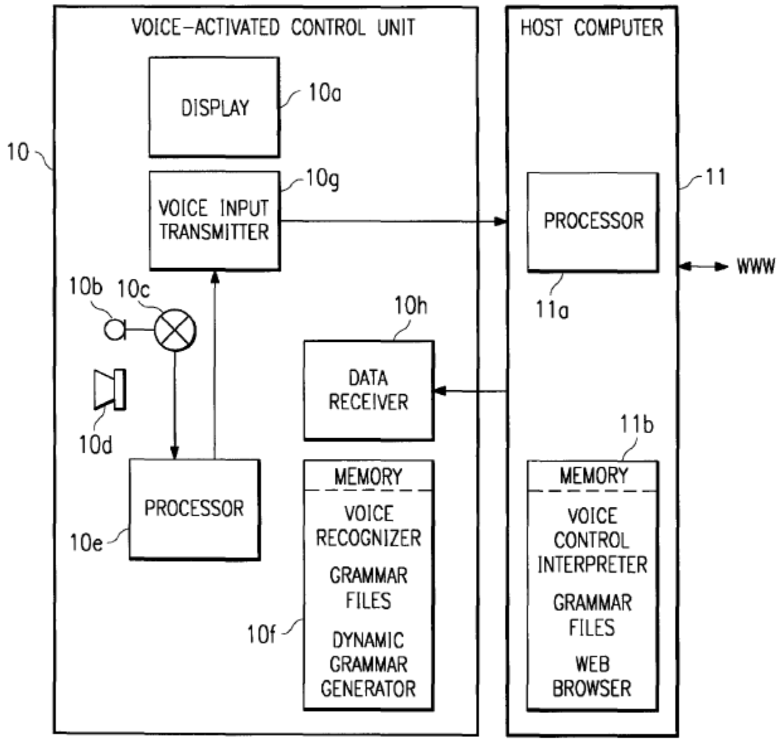


*FIG. 1*

(Ex. 1015, FIG. 1.)

82. A person of ordinary skill would further have recognized the benefits of implementing the device used in *Cheyer*'s process to be a remote control device for a television. For example, a person of ordinary skill would have recognized that configuring the device to be a portable remote control device for a television would have enabled the user to retrieve information via a broader set of devices, e.g., via a television as disclosed in *Thrift*. (Ex. 1015, 2:44-46.)

83. A person of ordinary skill would further have recognized that configuring a device to be a remote control device for a television would have been

a familiar, user-friendly configuration because remote controls for televisions were well-known long before the alleged invention of the ’718 patent. Implementing such a configuration would have been straightforward, because *Thrift*’s control unit 10 includes a wireless transmitter 10g and receiver 10h for remotely controlling and communicating with another device and a person of ordinary skill would have known how to program *Cheyer*’s handheld PDA, which similarly includes wireless communication components, to be a remote control for a television.

84. Furthermore, a person of ordinary skill would have recognized that configuring *Cheyer*’s mobile device to be a portable remote control device for a television would have been a predictable implementation, because it was well known at the time of the alleged invention of the ’718 patent to provide voice input to components for a television. For example, *Dureau*[3] discloses a system in which a “user can use a microphone or a telephone handset to provide voice data to the system,” whereby the “microphone may be connected to [a] set-top box, or it may be built into a remote control for the system,” and thereafter the “voice data is

---

[3] For claim 1, I am citing *Dureau* only to demonstrate knowledge of a person of ordinary skill.

transmitted to the server, which uses voice recognition software to convert the voice data into textual data." (Ex. 1016, 10:56-67.)

85. The above configuration would have been a mere combination of known components and technologies (e.g., *Cheyer*'s functionality relating to a voice interface for a device that remotely controls another device, and *Thrift*'s disclosure of a voice-controlled remote control device for a television), according to known methods (e.g., a person of ordinary skill knew how to program a device to implement wireless communication to remotely control a television), to obtain predictable results (e.g., a voice-controlled remote control device for a television that could be used to provide desired information to a user).

#### iii) [1.b] "(b) rendering an interpretation of the spoken request;"

86. *Cheyer* discloses this limitation. For instance, *Cheyer* discloses that a speech recognition agent recognizes a spoken English request and a "Natural Language (NL) Parser Agent" translates the request into the Interagent Communication Language (ICL). (Ex. 1012, 7 ("a speech recognition agent has been developed to provide transparent access to the Corona speech recognition system."), 9 ("Speech Recognition (SR) agent: The SR agent provides a mapping from the Interagent Communication Language to the API for the Decipher (Corona) speech recognition system . . . . Natural Language (NL) Parser Agent:

translates English expressions into the Interagent Communication Language (ICL)."), 9-10 (describing the NL agent, including parsing and semantic interpretation capabilities thereof), 11 ("The speech recognition agent monitors the status and results from its micro agent . . . . When the string is recognized, a translation is requested. . . . The English request is received by the NL agent and translated into ICL form."), FIG. 3 (reproduced below).)

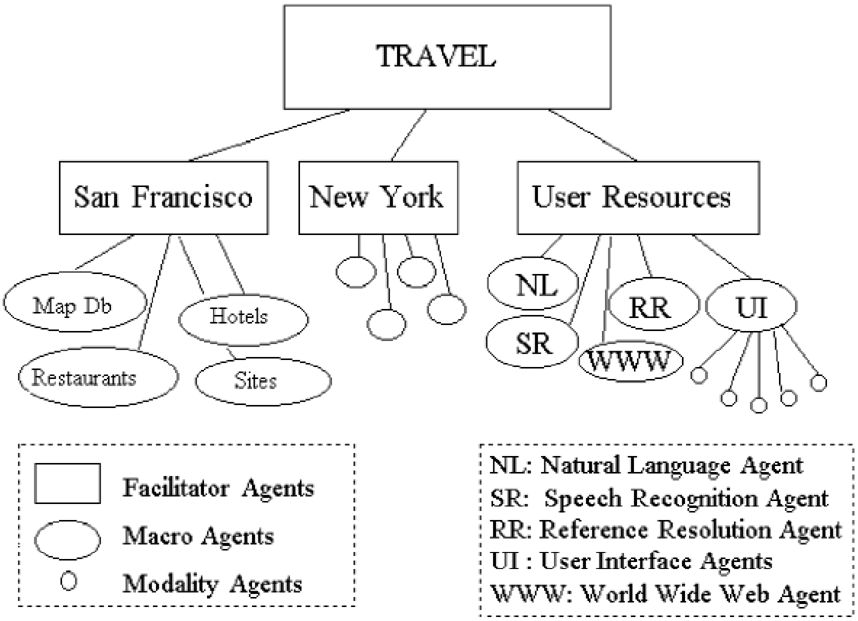


Figure 3: Agent Architecture for Map Application

(*Id.*, FIG. 3 (showing speech recognition agent and NL agent).) The "Decipher (Corona) speech recognition system" described in *Cheyer* (*id.*, 9) is just one example of a speech recognition system that was known before the alleged invention of the ’718 patent. (*See* above at Section V.A.) The use of a natural language (NL) parser as described in *Cheyer* was just an implementation of well-known technology, because it was well known before the alleged invention of the ’718 patent to perform parsing to extract meaning from phrases or sentences, e.g., phrases or sentences outputted by a speech recognizer. (*See* above at Section V.A.)

87. The speech recognition and ICL translation of the user’s speech input constitute an “interpretation of the spoken request,” so *Cheyer* discloses that the speech recognition agent and NL parser agent “render[] an interpretation of the spoken request.” In fact, the ’718 patent specification discloses the same use of a speech recognition agent and NL parser agent as disclosed in *Cheyer*. (*See, e.g.*, Ex. 1001, 14:33-36 (explaining that a “speech recognition agent 610” and “natural language (NL) agent 620” render an “interpretation in ICL format”); Ex. 1012, 7, 9-11.)

#### iv) [1.c] “(c) constructing a navigation query based upon the interpretation;”

88. *Cheyer* in combination with *Shwartz* discloses this limitation. For instance, *Cheyer* discloses that based on the interpretation provided by the speech recognition agent and NL parser agent, a domain agent “sends database requests”

asking for information related to the user's request, e.g., coordinates of items such as a reference or hotel. (Ex. 1012, 12.)

89. Therefore, *Cheyer* discloses a "navigation query" because *Cheyer*'s domain agent sends a database request ("navigation query") that enables the desired information to be retrieved for the user. (*See* above at Section VII.A.) *Cheyer*'s database request is a navigation query because it is an electronic query structured appropriately so as to navigate a data source of interest in search of desired information. (*See* above at Section VII.A; *see also* above at Section V.C regarding background information, known before the alleged invention of the '718 patent, regarding database queries for retrieving information from a database).)

90. While *Cheyer* may not expressly disclose "*constructing* a navigation query *based upon the interpretation*," a person of ordinary skill in the art would have been motivated in view of *Shwartz* to implement such features in *Cheyer*'s process. For example, while *Cheyer* discloses using database requests to retrieve information from a database to service a user's request (Ex. 1012, 11, 12; *see also id.*, 5, 6), *Cheyer* does not provide details regarding constructing such database requests or what they are based upon, but *Shwartz* discloses constructing a database query to navigate a database in search of desired information, as set forth below.

91. *Shwartz*, which is in the same technical field as *Cheyer* (e.g., natural language interface for servicing a user's request), discloses a "database retrieval system having a natural language interface" and further discloses that "[a] *database query is generated* . . . , enabling the retrieval and aggregation of data from [a] database to satisfy [a] natural language query." (Ex. 1013, Abstract (emphasis added).) For example, *Shwartz* discloses "retrieval of information from the application database in response to a query represented by the meaning representation." (*Id.*, 9:25-27.)

92. *Shwartz* explains that "[a] navigator and query language generator 38 is used to define optimal navigation paths through the database tables and columns to respond to the query, and to generate a meta-query language ('MQL')," and "[t]he metaquery language is used by a reporter and database access system 40 to *generate the code (e.g., structured query language ('SQL') code)* to actually retrieve the information from the application database." (*Id.*, 9:28-35 (emphasis added); *see also id.*, 7:19-22 ("*generation of the structured query language ('SQL') or other code* . . . to retrieve information from the database") (emphasis added), 17:1-19 (disclosing details regarding how to locate information from application database 32 responsive to a query).)

93. Thus, *Shwartz* teaches details of constructing a query suitable for retrieving, from a database (such as *Cheyer*'s remote databases), information desired by a user. A person of ordinary skill in the art would have understood *Shwartz* to teach constructing a "*navigation* query" because *Shwartz*'s foregoing generated query (e.g., SQL query) is an electronic query structured appropriately so as to navigate a particular data source in search of desired information. (*See* above at Section VII.A.) A person of ordinary skill would have had reason to look to *Shwartz* for implementing *Cheyer*'s process because both references pertain to obtaining information from a database. Such a person would have been motivated in view of *Shwartz* to configure *Cheyer*'s process to construct a database query so that information could be retrieved from a database in order to respond to the user's request. (*See* above at Section V.C for background information regarding database queries.)

94. Because *Cheyer*'s database request "ask[s] for the coordinates of the items in question" (e.g., the coordinates of the restaurant and the hotel referenced by the user's input query "How far is the restaurant from this hotel?") and the items in question are contained in the user's input query that is processed by the speech recognition agent and NL parser agent to interpret the meaning of the words in the input query, a person of ordinary skill in the art would have been motivated

to configure the combined *Cheyer-Shwartz* process to construct the database query *based upon the interpretation* that is rendered, similarly to the arrangement in *Shwartz*. (Ex. 1012, 11-12; Ex. 1013, 7:56-60 ("Software ('code') 302 is provided for use by natural language interface 34 to enable the production of the internal meaning representation 304."), 7:54-55 (disclosing a "query *interpretation* function" in connection with the natural language interface) (emphasis added), 9:20-35 ("*By accessing semantic and structural information* pertaining to an application database and residing in knowledge base 30, DBES 36 provides a retrieval specification that lists the tables and columns chosen, in accordance with column selection rules, for the retrieval of information from the application database in response to a query *represented by the meaning representation*. . . . The metaquery language is used by a reporter and database access system 40 to generate the code (e.g., structured query language ('SQL') code) to actually retrieve the information from the application database.") (emphases added), FIG. 1 (reproduced below and showing that SQL query is generated by reporter/database access system 40 based on interpretation of user's request rendered at natural language interface 34).)

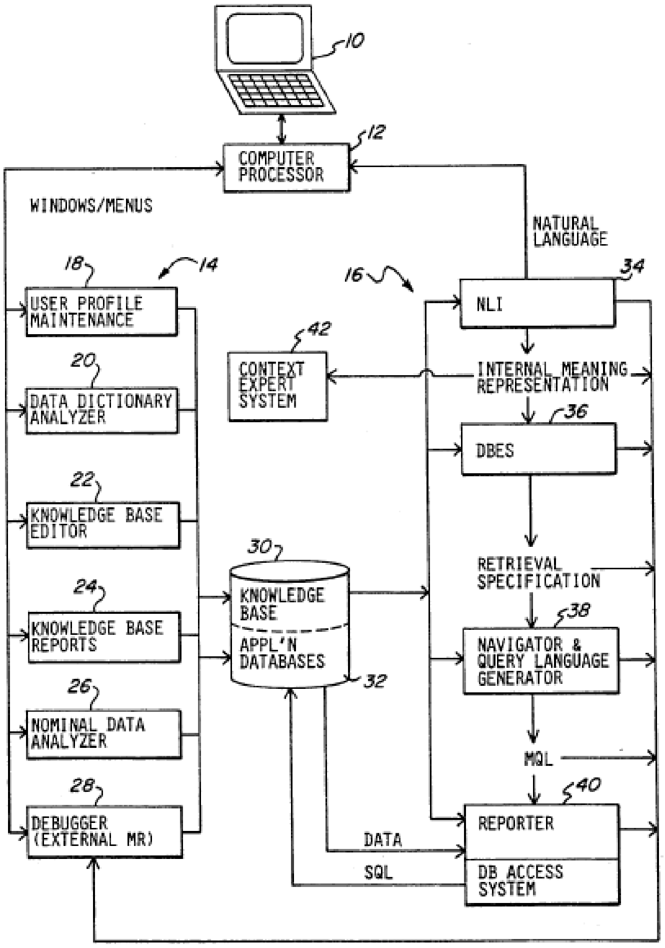


FIG. 1

(Ex. 1013, FIG. 1.)

95. In other words, a person of ordinary skill in the art would have been motivated to construct the database query in the combined *Cheyer*-*Shwartz* process based upon the interpretation of the user's spoken request so that the database query could properly specify information to be retrieved from *Cheyer*'s remote database.

96. A person of ordinary skill in the art would have been capable of implementing the above configuration for the combined *Cheyer-Shwartz* process, and would have had a reasonable expectation of success regarding the outcome, particularly because *Shwartz* is directed to a system for processing natural language requests from a user, like *Cheyer*. (Ex. 1013, Abstract; Ex. 1012, 1 (at Abstract).) This would have been a straightforward implementation that merely involved constructing a navigation query to access a database in a predictable manner. Such an implementation would have been a mere combination of elements and technologies (e.g., a database request for servicing a query, as taught by *Cheyer*, and construction of a database query, i.e., database request, as taught by *Shwartz*), according to known methods (e.g., *Shwartz* describes how to construct the query, and *Cheyer* describes its role in a system for servicing a user's request), to provide predictable results (e.g., retrieving information desired by the user from a database).

### v) [1.d] "(d) utilizing the navigation query to select a portion of the electronic data source; and"

97. *Cheyer* alone and/or in combination with *Shwartz*[4] discloses this limitation. For instance, *Cheyer* discloses that a database agent utilizes the navigation query to retrieve from a database information requested by a user ("select a portion of the electronic data source"). *Cheyer* discloses various examples of such "portion[s] of the electronic data source," such as "maps for each city, as well as icons, vocabulary and information about available hotels, restaurants, movies, theaters, municipal buildings and tourist attractions" (Ex. 1012, 10), "the French restaurants within 1 mile of this hotel" (*id.*, 5) or "all available information about Alcatraz" (*id.*, 5).

98. *Cheyer* discloses that a type of agent called a "facilitator" routes information to agents in the Open Agent Architecture. (*See id.*, 7 ("Facilitators provide content-based message routing, global data management, and process

---

[4] As discussed above for limitation [1.c], a person of ordinary skill would have been motivated in view of *Shwartz* to modify *Cheyer*'s process to construct a "navigation query." A person of ordinary skill would also have been motivated to configure the combined *Cheyer*-*Shwartz* process to implement the features relating to "navigation query" in limitation [1.d] and claim 4.

coordination for their set of connected agents.”), 8 (“when queries arrive in Interagent Communication Language form, [facilitators] are responsible for breaking apart any complex queries and for distributing goals to the appropriate agents”); *see also id.*, 9 (“facilitator agent”).)

99. *Cheyer* discloses that database agents provide information (e.g., about maps, places of interest, movies, etc.) relevant to the user’s request. (Ex. 1012, 10 (“database agents provide maps for each city, as well as icons, vocabulary and information about available hotels …”).)

100. *Cheyer*’s database agents retrieve information from a database based on database requests. (Ex. 1012, 10 (“a domain agent will try to resolve the definite reference by sending database agent requests”).) Thus, when a database request is constructed for retrieving information from a database in response to a user’s input such as “Display the French restaurants within 1 mile of this hotel” (*id.*, 5) or “Show me all available information about Alcatraz” (*id.*), a corresponding database request is routed to a database agent that services the request by utilizing the database request (“navigation query”) to access the database.

101. While *Cheyer* discloses “access to existing data sources” (Ex. 1012, 1 (at Abstract)), “access to various heterogeneous data and knowledge sources” (*id.*,

7), "access [to] a wide variety of data sources, including information stored in HTML form on the World Wide Web" (*id.*, 4), and various types of databases, including "Prolog databases, X-500 hierarchical databases, and data loaded automatically by scanning HTML pages from the World Wide Web (WWW)" (*id.*, 10), *Cheyer* does not expressly disclose that the database agent "select[s] a portion" of the disclosed electronic data source. However, a person of ordinary skill in the art would have understood that *Cheyer* necessarily discloses that feature. Such a skilled person would have had this understanding because "database requests" (*id.*, 12) were well known to be for retrieving or selecting a portion of a database. If a portion of the database that contains the "maps for each city" or "information about available hotels . . . and tourist attractions" (*id.*, 10) were not selected by *Cheyer*'s database agent, then the database agent would not have been able to provide the information that the user requested, such as "the French restaurants within 1 mile of this hotel" (*id.*, 5) or "all available information about Alcatraz" (*id.*).

102. I have been asked to assume that *Cheyer* does not disclose "select[ing] a portion of the electronic data source." Under that assumption, it is my opinion that a person of ordinary skill would have been motivated in view of *Cheyer* and *Shwartz* to implement this feature in *Cheyer*'s process. *Shwartz* discloses

"retrieval and aggregation of data from [a] database to satisfy [a] natural language query" (Ex. 1013, Abstract) and "identify[ing] an optimal set of database elements to satisfy the query" (*id.*, 17:10-11), e.g., by choosing particular "tables and columns" (*id.*, 9:24-27). Additionally, *Shwartz* discloses "generat[ing] . . . code (e.g., structured query language ('SQL') code) to actually retrieve the information from the application database" (*id.*, 9:33-35), and a person of ordinary skill in the art would have understood that SQL code (e.g., a SELECT statement in SQL code) was intended to select a portion of a database. (*See* above at Section V.C; *see also* Ex. 1013, 7:19-22 ("generation of the structured query language ('SQL') or other code that is ultimately produced by the query system to retrieve information from the database").)

103. A person of ordinary skill in the art would have been motivated, in light of the teachings of *Cheyer* and *Shwartz*, to configure *Cheyer*'s process to select a portion of any of the databases disclosed by *Cheyer*. Such a skilled person would have recognized that selecting a portion of a database responsive to the user's request would have enabled the combined *Cheyer*-*Shwartz* process and system to provide desired information to the user. This would have been a straightforward configuration, because it would have been merely a combination of known components and technologies (e.g., *Cheyer*'s database and database

requests, and *Shwartz*'s "structured query language ('SQL') or other code" for retrieving a portion of a database (Ex. 1013, 7:19-22)), according to known methods (e.g., retrieving information from a database using database requests), to obtain predictable results (selecting a portion of a database in response to a database request). (*See* above at Section V.C regarding knowledge of one of skill in the art regarding programming a computer system to retrieve information from a database.)

104. A person of ordinary skill would have recognized that an alternative to selecting a *portion* of a database would have been to select the entire database for downloading. A person of ordinary skill would have considered such an alternative to be resource-expensive and/or wasteful in many scenarios (e.g., in the scenario of a large database and a user's request that could be serviced by using only a portion of the database). Therefore, a person of ordinary skill would have been motivated to select a portion of the database as discussed above.

#### vi) [1.e] "(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user."

105. *Cheyer* alone and/or in combination with *Shwartz* discloses this limitation. For instance, in the examples of "Display the French restaurants within 1 mile of this hotel" (Ex. 1012, 5) or "Show me all available information about

Alcatraz" (*id.*), *Cheyer* discloses displaying the French restaurants within 1 mile of the hotel specified by the user or displaying all available information about Alcatraz. (*See id.*, 10 ("Interface Agent: This macro agent [that] is responsible for managing what is currently being displayed to the user …."), 11-12 ("[T]he domain agent (RR) … then calculates the distance according to the scale of the currently displayed map, and requests the user interface to produce output displaying the result of the calculation.").) *Cheyer* also discloses "access to existing data sources including the World Wide Web." (*Id.*, Abstract.)

106. Based on the foregoing disclosures, a person of ordinary skill would have understood that *Cheyer* necessarily discloses transmitting the selected portion of the electronic data source from the remote server at which such data sources are located ("the network server") to the user's mobile computing device ("the mobile information appliance of the user"). For example, if such data were not transmitted from the remote server to the user's mobile device, the user could not have obtained the information that he/she desired. Indeed, *transmitting* data from a remote server to a user's computing device was known to be a necessary aspect of data communications involving the Web, which *Cheyer* discloses. (Ex. 1012, Abstract ("access to existing data sources including the World Wide Web.").)

107. I have been asked also to assume *Cheyer* does not disclose transmitting the selected portion of the electronic data source from the remote server to the user's mobile device. Under that assumption, it is my opinion that a person of ordinary skill would have been motivated in view of *Shwartz* to implement such features. As discussed above for limitation [1.d], a person of ordinary skill would have been motivated in view of *Shwartz* to select a portion of the electronic data source (*see* above at Section IX.A.1.v), and in view of *Shwartz*'s disclosure of displaying retrieved data on a user's computer a person of ordinary skill would further have been motivated to configure the combined process to transmit the selected portion from the remote network server to the user's mobile device. (Ex. 1013, 5:9-11 ("Data retrieved from a database in response to a natural language query can be displayed on a user's workstation.").) A person of ordinary skill would have known how to implement data communications involving the Web, which *Cheyer* discloses (Ex. 1012, Abstract), and would have been motivated to implement such transmitting in order to achieve a working application as disclosed in *Cheyer*. (*See* above at Section V.C.) Indeed, a person of ordinary skill would not only have been motivated but would have naturally expected to configure *Cheyer*'s process to transmit the selected portion of the electronic data source from the remote server to the user's mobile device, in

order to achieve *Cheyer*'s objective of enabling a user "to transparently access a wide variety of data sources, including information stored in HTML form on the World Wide Web." (Ex. 1012, 4.)

108. A person of ordinary skill would have recognized that transmitting an entire database instead of transmitting a selected portion of the database would have presented challenges in terms of network resources and time in many scenarios (e.g., particularly in the example of a large database), and therefore, would have been motivated to transmit the selected portion of the database in order to avoid or mitigate such challenges.

109. This would have been a mere combination of known components and technologies (e.g., *Cheyer*'s disclosure of an application that retrieves information from a remote data source such as one located on the Web, *Cheyer*'s disclosure of a PDA that a person of ordinary skill would have known was capable of receiving information transmitted by a remote server, and *Shwartz*'s disclosure of displaying retrieved data on a user's computer), according to known methods (e.g., implementing data communications involving the Web in a known manner), to obtain predictable results (e.g., sending information from a remote server to the user's mobile device). (*See* above at Section V.C.)

### 2. Claims 2 and 3

#### i) [2.a]/[3.a] "The method of claim 1, wherein the step of rendering the interpretation of the spoken request is performed by the mobile information appliance."

110. While *Cheyer* discloses a "server machine which will manage . . . natural language processing and speech recognition for the application" (Ex. 1012, 6), a person of ordinary skill would have been motivated in view of *Thrift* to configure the combined *Cheyer-Thrift-Shwartz* process to perform the speech recognition and natural language processing ("the step of rendering the interpretation of the spoken request") at the user's mobile computing device ("the mobile information appliance").

111. *Cheyer* discloses "[t]he user interface must be light and fast enough to run on a handheld PDA while able to access applications and data that *may* require a more powerful machine" (Ex. 1012, 4), which suggests that in some situations (e.g., when the user's handheld PDA is sufficiently powerful) a more powerful machine (e.g., server remote from the PDA) may not be needed. A person of ordinary skill would have understood *Cheyer*'s foregoing disclosure as providing guidance as to when a remote server for performing speech recognition and natural language processing would or would not be appropriate (i.e., the resource capabilities of the PDA are central to this issue).

112. *Thrift*, in the same technical field as *Cheyer* (e.g., providing information to a user based on voice input), discloses a client-server architecture in the speech processing context but also explains that in some instances a host computer 11 (the server in *Thrift*'s client-server architecture) is not needed for at least some speech processing tasks. (Ex. 1015, 3:1-24 ("In the embodiment of FIG. 1, control unit 10 performs *all or part* of the voice recognition process and delivers speech data to host computer 11 via transmitter 10g … . [I]n its simplest form control unit would transmit audio data directly from microphone 10b to host system 11, which would perform all processing. In the case where control unit 10 performs *all or part* of the voice recognition process, control unit 10 has a processor 10e…. If control unit performs only some voice processing, it may perform one or more of the 'front end' processes …. If control unit 10 performs all voice recognition processes, memory 10f stores these processes (as a voice recognizer) as well as grammar files.") (emphases added).)

113. Thus, *Thrift* indicates that it was known before the alleged invention of the '718 patent that tasks could either be allocated to a separate server or performed at the client, depending on particular system needs.

114. A person of ordinary skill would have understood *Thrift*'s disclosure regarding control unit 10 (the client in *Thrift*'s client-server architecture)

performing all or part of a *voice recognition* process to also be applicable to modifying *Cheyer*'s process to have the user's PDA perform all or part of speech recognition *and natural language processing*, because a person of ordinary skill would have understood that *Thrift*'s foregoing disclosure is relevant to allocation of tasks in a variety of computational contexts. In other words, it would have been useful to assign natural language processing to the user's PDA, because natural language processing, like speech recognition, was a task that involved processing data.

115. A person of ordinary skill would have had reason to consider the teachings of *Thrift* (in the same technical field as *Cheyer*) when implementing *Cheyer*'s process and would have seen that *Thrift* discloses that certain tasks may be assigned to either control unit 10 or to host system 11. A person of ordinary skill would have understood that *Thrift*'s disclosure of control unit 10 performing "all or part" of a voice recognition process (Ex. 1015, 3:1-2, 3:9-10) meant that the choice of which tasks to allocate to the control unit 10 as opposed to host system 11 was determined by system implementation details such as relative resource capabilities. Based on *Thrift*'s disclosure of "control unit 10 perform[ing] *all* voice recognition processes" in one scenario, a person of ordinary skill would have recognized the possibility and value of configuring *Cheyer*'s PDA to perform the

speech recognition and natural language processing functions disclosed in *Cheyer*. (Ex. 1015, 3:22-23 (emphasis added).)

116. For example, a person of ordinary skill would have been motivated to make the above modification in order to reduce communications latency, e.g., by eliminating communications to and from a remote server regarding speech recognition and natural language processing. A person of ordinary skill would also have been motivated to make this modification to simplify the architecture of *Cheyer*'s system, because with the functions of speech recognition and natural language processing performed at the PDA then a separate speech server would not have been needed for such processing. A person of ordinary skill would have been capable of making this modification, as the choice of a single computer design or a client-server design was a mere choice among a finite number of known alternatives with predictable outcomes.

### 3. Claim 4

#### i) [4.a] "The method of claim 1, further comprising the steps of soliciting additional input from the user, including user interaction in a modality different than the original request;"

117. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation. For instance, *Cheyer* discloses soliciting additional input from the user,

including user interaction via pen (e.g., gestures and/or handwriting using a pen) ("a modality different than the original request").

118. At the outset, *Cheyer* discloses several examples in which the user provides additional input beyond just spoken input, including user interaction in a modality different than the original spoken request. (*See, e.g.*, Ex. 1012, 5-6 ("the user is presented with a *pen sensitive map display on which drawn gestures and written natural language statements* may be combined with spoken input. . . . During input, requests can be entered using *gestures* (Figure 2), *handwriting*, voice, or a combination of *pen* and voice. . . . For gestural commands, which are handled locally on the user interface machine, a response is produced in less than one second.") (emphases added), 6 ("synergistic combination of pen and voice, by [the user] speaking 'What is the distance from here to this hotel?' while *simultaneously indicating the specified locations by pointing or circling*") (emphasis added); *see also* above at Section V.B regarding background information, known before the alleged invention of the '718 patent, regarding multimodal input.)

119. *Cheyer* further discloses prompting the user for additional input ("*soliciting* additional input from the user"). For example, *Cheyer* explains circumstances in which additional input may be solicited from the user, such as

when a user's original request is ambiguous or underspecified. (*See, e.g.*, Ex. 1012, 6 ("[I]n our example of synergistic combination of pen and voice, the arguments to the verb 'distance' can be specified before, at the same time, or shortly after the vocalization of the request to calculate the distance. *If a user's request is ambiguous or underspecified, the system will wait several seconds and then issue a prompt requesting additional information*.") (emphasis added).) For example, *Cheyer* discloses prompting the user for an indication (e.g., via a gesture) as to what the user means by the phrase "the hotel" in the user's spoken request. (*Id.*, 11 ("An important task for the interface agent is to record which objects of each type are currently salient, in order to resolve contextual references such as 'the hotel' or 'where I was before.' Deictic references are resolved by *gestural or direct manipulation commands*. If no such indication is currently specified, the user interface agent waits long enough to give the user an opportunity to supply the value, and then *prompts the user for it*.") (emphases added), 12 ("The interface agent . . . waits for the user to make a *gesture* indicating '[this] hotel', *issuing prompts if necessary*.") (emphasis added).)

### ii) [4.b] "refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source."

120. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation. For instance, *Cheyer* discloses that in the example of a user input "How far is the restaurant from this hotel" (Ex. 1012, 11), the database request ("navigation query") is refined based upon "a gesture indicating '[this] hotel'" (*id.*, 12), because there is an ambiguity regarding what "this hotel" refers to. *Cheyer* discloses that a "reference resolution agent (RR) . . . asks for resolution of" a reference such as "[this] hotel" and that "[w]hen the references have been resolved, the domain agent . . . sends database requests . . . ." (*Id.*) Thus, *Cheyer* discloses that the database request is refined based upon the additional input from the user that clarifies what the user means by "this hotel", and the domain agent sends the refined database request after the ambiguity regarding the reference "this hotel" (*id.*, 11) has been resolved (*id.*, 12). *Cheyer* discloses using the refined database request ("refined navigation query") to retrieve from the remote database location information regarding the hotel specified by the user ("to select a portion of the electronic data source"), so that the distance requested by the user can be calculated. (*Id.*, 10 (describing details of database agent).)

121. As another example, *Cheyer* discloses that the user may speak "Display the French restaurants within 1 mile of this hotel." (Ex. 1012, 5.) The phrase "this hotel" in this example's spoken query, which is similar to the above-described example involving "the hotel" at page 11 of *Cheyer*, is ambiguous and requires clarification. After the user provides such additional input so that the ambiguity can be resolved, *Cheyer*'s database agent uses a refined database query that takes into account the additional information regarding the identity of the hotel ("the refined navigation query") to select a portion of a database containing maps or "information about available restaurants" relevant to the user's query ("a portion of the electronic data source"). (*Id.*, 10; *see also id.*, 11 ("resolve contextual references such as 'the hotel' . . . by gestural or direct manipulation commands."); *see also* above at Section V.C regarding background information, known before the alleged invention of the '718 patent, regarding retrieving information from a database.)

122. I have been asked also to assume *Cheyer* does not disclose the feature "to select a portion of the electronic data source." Under that assumption, it is my opinion that a person of ordinary skill would have been motivated in view of *Shwartz* to implement that feature in the combined *Cheyer-Shwartz* process for at

least the same reasons discussed above for limitation [1.d]. (*See* above at Section IX.A.1.v.)

### 4. Claim 6

#### i) [6.a] "The method of claim 1, wherein steps (a)-(d) are performed with respect to multiple users."

123. *Cheyer* in combination with *Thrift* and *Shwartz* discloses or suggests this limitation. For example, *Cheyer* discloses an application including a user interface that runs on a handheld PDA or a PC (Ex. 1012, Abstract, 4, 6) and further discloses multiple users (*id.*, 1-2 (referring to multiple "users")). A person of ordinary skill would have understood that when a plurality of simultaneous users using respective PDAs run *Cheyer*'s application, the method of claim 1, including steps (a)-(d) recited therein, is necessarily performed with respect to multiple users. Even if this were not the case, a person of ordinary skill would have been motivated to perform steps (a)-(d) with respect to multiple users, e.g., to enable a wider range of people than just one person to be able to use the combined *Cheyer*-*Thrift*-*Shwartz* process. A person of ordinary skill would have recognized that enabling multiple users to use the combined process would have beneficial, e.g., in order to provide information to more people. A person of ordinary skill would have been motivated to implement such a feature particularly because *Cheyer* discloses access to databases on the Web and prior to the alleged invention

of the ’718 patent the Web involved providing multiple users with access to websites. (Ex. 1012, Abstract, 10; *see also* above at Section V.C.)

### 5. Claims 8, 9

#### i) [8.a] “The method of claim 1, wherein the mobile information appliance is a portable computing device.”

#### ii) [9.a] “The method of claim 8, wherein the portable computing device is a personal digital assistant.”

124. *Cheyer* combined with *Thrift* and *Shwartz* discloses these limitations. *Cheyer* discloses that the application discussed above for claim 1 runs on a handheld personal digital assistant (PDA), which a person of ordinary skill would have understood to be a portable computing device. (Ex. 1012, 4 (“The user interface must be light and fast enough to run on a handheld PDA … .”), 6 (“The user interface runs on … a Dauphin *handheld PDA* … . The result is a *mobile system* that provides a synergistic pen/voice interface to remote databases.”) (emphasis added); *see also* Ex. 1012, Abstract (“The application is distinguished by a synergistic combination of handwriting, gesture and speech modalities; access to existing data sources including the World Wide Web and a *mobile handheld interface*.”) (emphasis added), 12 (“mobile, synergistic pen/voice interface”).)

125. A person of ordinary skill would have recognized that the remote control device (“mobile information appliance”) in the combined *Cheyer-Thrift-*

*Shwartz* process (discussed above for claim 1)could have additionally been a portable computing device (e.g., PDA), and would have been motivated to implement the device to be both a remote control device and a portable computing device (e.g., PDA). For example, a person of ordinary skill would have recognized that the attributes of a remote control device and of a portable computing device (e.g., PDA) were not mutually exclusive, and that these were separate features that could have beneficially have been co-implemented. Indeed, a person of ordinary skill would have been motivated to co-implement both of these features in order to provide a richer feature set for users and to enable a user to perform remote control functionality with an existing device such as his/her portable computing device, e.g., PDA. Such an implementation would have promoted efficiency, e.g., by using a single device to perform multiple features, and would have been consistent with the knowledge of a person of ordinary skill and the expectations of consumers regarding multi-function devices.

126. An article by Konstan published in 1994 ("*Konstan*") shows that before the alleged invention of the '718 patent it was known to implement a mobile device that was both a PDA (which was a known type of portable computing device) and a remote control for a television. For example, *Konstan* discloses that "the emergence of personal digital assistants has created new possibilities for

programmed device control. (Ex. 1033, 812.) *Konstan* further discloses that "[b]asic PDA's can dial stored phone numbers [and] [m]ore advanced ones can also … store and play back infrared control sequences such as are used for controlling televisions and other consumer audio/video devices." (*Id.*; *see also id.* ("personal digital assistants … are now capable of learning and generating control sequences to control a wide range of devices").) Therefore, *Konstan* demonstrates that a person of ordinary skill would have known how to, and would have been motivated to, make the above implementation.

#### 6. Claim 10

##### i) "A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:"

127. I have been asked to assume that the preamble of claim 10 is limiting. Under that assumption, it is my opinion that *Cheyer* discloses the limitations therein for at least the same reasons as presented above regarding the preamble of claim 1. (*See* above at Section IX.A.1.i for citations and analysis regarding preamble of claim 1; *see also* below at Sections IX.A.6.ii-vi for the remaining limitations of this claim.)

128. *Cheyer* discloses an "application" (Ex. 1012, 1-9, 11-12) that "runs on pen-equipped PC's or a Dauphin handheld PDA" (*id.*, 6). *Cheyer* discloses that "[t]o implement the described application, a distributed network of heterogeneous *software* agents was augmented by appropriate functionality for developing synergistic multimodal applications." (*Id.*, 1 (emphasis added).) Therefore, a person of ordinary skill would have understood that *Cheyer* discloses a "computer program embodied on a computer readable medium" as claimed.

#### ii) [10.a] "(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;"

129. *Cheyer* in combination with *Thrift* discloses this limitation for at least the same reasons as presented above regarding limitation [1.a]. (*See* above at Section IX.A.1.ii for citations and analysis regarding limitation [1.a].)

130. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.a], and thus *Cheyer* discloses a "code segment" as

in limitation [10.a]. (*See* above at Section VII.B.) Even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

#### iii) [10.b] "(b) a code segment that renders an interpretation of the spoken request;"

131. *Cheyer* discloses this limitation for at least the same reasons as presented above regarding limitation [1.b]. (*See* above at Section IX.A.1.iii for citations and analysis regarding limitation [1.b].)

132. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.b], and thus *Cheyer* discloses a "code segment" as in limitation [10.b]. (*See* above at Section VII.B.) Even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a

processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

#### iv) [10.c] "(c) a code segment that constructs a navigation query based upon the interpretation;"

133. *Cheyer* in combination with *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [1.c]. (*See* above at Section IX.A.1.iv for citations and analysis regarding limitation [1.c].)

134. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.c], and thus *Cheyer* discloses a "code segment" as in limitation [10.c]. (*See* above at Section VII.B.) Even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013,

4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

135. Although *Cheyer* does not expressly describe in detail the limitation "constructs a navigation query," a person of ordinary skill in the art would have been motivated in view of *Shwartz* to implement that feature in *Cheyer*'s computer program, for at least the same reasons as discussed above for limitation [1.c]. (*See* above at Section IX.A.1.iv.)

#### v) [10.d] "(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and"

136. *Cheyer* in combination with *Shwartz*[5] discloses this limitation for at least the same reasons as presented above regarding limitation [1.d]. (*See* above at Section IX.A.1.v.)

---

[5] As discussed above for limitation [10.c], a person of ordinary skill would have been motivated in view of *Shwartz* to modify *Cheyer*'s computer program to construct a "navigation query." A person of ordinary skill would also have been motivated to configure the combined *Cheyer*-*Shwartz* computer program to implement the "navigation query" feature in limitation [10.d] and claim 13.

137. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.d], and thus *Cheyer* discloses a "code segment" as in limitation [10.d]. (*See* above at Section VII.B.) Even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

138. I have been asked also to assume that *Cheyer* does not disclose "select a portion of the electronic data source." Under that assumption, it is my opinion that a person of ordinary skill in the art would have been motivated in view of the combined teachings of *Cheyer* and *Shwartz* to implement this feature in *Cheyer*'s computer program, for at least the same reasons as discussed above for limitation [1.d]. (*See* above at Section IX.A.1.v.)

#### vi) [10.e] "(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user."

139. *Cheyer* discloses this limitation for at least the same reasons as presented above regarding limitation [1.e]. (*See* above at Section IX.A.1.vi.)

140. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.e], and thus *Cheyer* discloses a "code segment" as in limitation [10.e]. (*See* above at Section VII.B.) Even if *Cheyer* were found not to provide for such an implementation, as recognized by *Shwartz* the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

### 7. Claim 12

#### i) [12.a] "The computer program of claim 10, wherein the rendering of the interpretation of the spoken request is performed by the mobile information appliance."

141. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claim 3. (*See* above at Section IX.A.2.)

### 8. Claim 13

#### i) [13.a] "The computer program of claim 10, further comprising a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;"

142. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.a]. (*See* above at Section IX.A.3.i for citations and analysis regarding limitation [4.a].)

143. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [13.a], and thus *Cheyer* discloses a "code segment" as

in limitation [13.a]. (*See* above at Section VII.B.) Even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

#### ii) [13.b] "a code segment that refines the navigation query, based upon the additional input; and a code segment that uses the refined navigation query to select a portion of the electronic data source."

144. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.b]. (*See* above at Section IX.A.3.ii for citations and analysis regarding limitation [4.b].)

145. A person of ordinary skill in the art would have understood based on *Cheyer*'s disclosure of an "application" (Ex. 1012, 1-9, 11-12) that runs on a PC or PDA (*id.*, 6) and further based on *Cheyer*'s disclosure of software agents (*id.*, 1) that *Cheyer*'s application includes software running on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [10.b], and thus *Cheyer* discloses a "code segment" as

in limitation [10.b]. (*See* above at Section VII.C.) Even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement software code was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

### 9. Claim 15

#### i) [15.a] "The computer program of claim 10, wherein code segments (a)-(d) are executed with respect to multiple users."

146. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claim 6. (*See* above at Section IX.A.4.)

### 10. Claims 17, 18

#### i) [17.a] "The computer program of claim 10, wherein the mobile information appliance is a portable computing device."

#### ii) [18.a] "The computer program of claim 17, wherein the portable computing device is a personal digital assistant."

147. *Cheyer* in combination with *Thrift* and *Shwartz* discloses these limitations for at least the same reasons as presented above regarding claims 8 and 9. (*See* above at Sections IX.A.5.i-ii.)

### 11. Claim 19

#### i) "A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:"

148. I have been asked to assume that the preamble of claim 19 is limiting. Under that assumption, it is my opinion that *Cheyer* discloses the limitations therein for at least the same reasons as presented above regarding the preamble of claim 1. (*See* above at Section IX.A.1.i for citations and analysis above regarding preamble of claim 1; *see also* below at Section IX.A.11.ii-vi for the remaining limitations of this claim.)

149. In addition to disclosing a "method" as recited in claim 1, *Cheyer* discloses a "system" utilizing an application that runs on a PC or PDA, and thus discloses a "system" as recited in the preamble of claim 13. (*See* above at Section

IX.A.1.i; Ex. 1012, 6; *see also id.*, 4 ("our *system* produces a richer mixing of modalities by adding both gestural and written language as input modalities . . . . When designing the architecture for the *system*, other criteria were considered as well . . . . The map functionality, interface design, and classes of input data of the *system* presented here is based on a design by Oviatt and Cohen . . . .") (emphases added), 6 ("If a user's request is ambiguous or underspecified, the *system* will wait several seconds and then issue a prompt requesting additional information. . . . The result is a mobile *system* that provides a synergistic pen/ voice interface to remote databases. . . . In general, the speed of the *system* is quite acceptable.") (emphases added), 12 ("The *system* described here is one of the first that accepts commands made of synergistic combinations of spoken language, handwriting and gestural input.") (emphases added).)

#### ii) [19.a] "(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;"

150. *Cheyer* in combination with *Thrift* discloses this limitation for at least the same reasons as presented above regarding the preamble of claim 1 and limitation [1.a]. (*See* above at Sections IX.A.1.i-ii.) For the reasons presented above regarding limitation [1.a], a person of ordinary skill would have been

motivated to implement in the combined system a mobile device comprising a portable remote control device or a set top box for a television, and such a person would have been motivated to configure the mobile device to be *operable to* receive a spoken request for desired information from the user, so that the user could use the mobile device to retrieve information via voice input as disclosed by *Cheyer*.

#### iii) [19.b] "(b) spoken language processing logic, operable to render an interpretation of the spoken request;"

151. *Cheyer* discloses this limitation for at least the same reasons as presented above regarding limitation [1.b]. (*See* above at Section IX.A.1.iii.)

152. Additionally, because *Cheyer* discloses an application implemented in software (*see* above at Section IX.A.6.i) and a person of ordinary skill would have understood that such software runs on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [19.b], and for the reasons discussed above for limitation [1.b], *Cheyer* discloses a "spoken language processing logic, operable to" perform the functionality recited in limitation [19.b]. (*See* above at Sections VII.C, IX.A.1.iii, IX.A.6.iii.) Indeed, even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement logic was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013,

4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("computer processor 12 … controls the overall operation of the system").)

#### iv) [19.c] "(c) query construction logic, operable to construct a navigation query based upon the interpretation;"

153. *Cheyer* in combination with *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [1.c]. (*See* above at Section IX.A.1.iv.)

154. Additionally, because *Cheyer* discloses an application implemented in software (*see* above at Section IX.A.6.i) and a person of ordinary skill would have understood that such software runs on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [19.c], and for the reasons discussed above for limitation [1.b], *Cheyer* discloses a "query construction logic, operable to" perform the functionality recited in limitation [19.c]. (*See* above at Sections VII.C, IX.A.1.iii, IX.A.6.iii.) Indeed, even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement logic was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by

executing the processor), 6:29-30 ("computer processor 12 … controls the overall operation of the system").)

155. Although *Cheyer* does not expressly describe in detail the limitation "construct a navigation query," a person of ordinary skill would have been motivated in view of *Shwartz* to implement that feature in *Cheyer*'s system, for at least the same reasons as discussed above for limitation [1.c]. (*See* above at Section IX.A.1.iv.)

#### v) [19.d] "(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and"

156. *Cheyer* in combination with *Shwartz*[6] discloses this limitation for at least the same reasons as presented above regarding limitation [1.d]. (*See* above at Section IX.A.1.v.)

---

[6] As discussed above for limitation [19.c], a person of ordinary skill would have been motivated in view of *Shwartz* to modify *Cheyer*'s system to construct a "navigation query." A person of ordinary skill would further have been motivated to configure the combined *Cheyer*-*Shwartz* system to implement the "navigation query" feature in limitation [19.d] and claim 22.

157. Additionally, because *Cheyer* discloses an application implemented in software (*see* above at Section IX.A.6.i) and a person of ordinary skill would have understood that such software runs on a microprocessor configured to perform various functionalities, including the functionality corresponding to limitation [19.d], and for the reasons discussed above for limitation [1.b], *Cheyer* discloses a "navigation logic, operable to" perform the functionality recited in limitation [19.d]. (*See* above at Sections VII.C, IX.A.1.iii, IX.A.6.iii.) Indeed, even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement logic was routine and commonplace at the time of the alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("computer processor 12 … controls the overall operation of the system").)

#### vi) [19.e] "(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user."

158. *Cheyer* alone and/or in combination with *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [1.e]. (*See* above at Section IX.A.1.vi.) A person of ordinary skill would have understood that *Cheyer* alone and/or in combination with *Shwartz* necessarily

discloses an *electronic communications infrastructure* for performing the transmitting of limitation [19.e]. A person of ordinary skill would have had this understanding because without an electronic communications infrastructure, a system like that disclosed in *Cheyer* and *Shwartz*, which involve retrieving information from a remote system (*see* above at Sections IX.A.i, vi, IX.11.i) would not have been possible. Indeed, an electronic communications infrastructure was a necessary component of a remote server (e.g., web server) such as disclosed by *Cheyer* in the context of a Web-based data source. (*See* above at Section IX.A.1.i; Ex. 1012, Abstract ("access to existing data sources including the World Wide Web"), 6 ("a mobile system that provides a synergistic pen/voice interface to remote databases"); *see also* above at Section V.C.)

159. I have been asked also to consider a scenario in which the claimed "electronic communications infrastructure for transmitting . . ." requires software running on a microprocessor configured to perform transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. In that scenario, it is my opinion that *Cheyer* discloses this limitation for at least the same reasons discussed above regarding limitation [10.e]. (*See* above at Section IX.A.6.vi for citations and analysis regarding limitation [10.e].)

### 12. Claim 21

#### i) [21.a] "The system of claim 19, wherein the spoken language processing logic renders the interpretation of the spoken request at the mobile information appliance."

160. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claim 3. (*See* above at Section IX.A.2.)

### 13. Claim 22

#### i) [22.a] "The system of claim 19, further comprising user interaction logic operable to solicit additional input from the user, including user interaction in a modality different than the original request; and"

161. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.a]. (*See* above at Section IX.A.3.i.)

162. Additionally, because *Cheyer* discloses an application implemented in software (*see* above at Section IX.A.6.i for citations and analysis regarding the preamble of claim 10), and for the reasons discussed above for limitation [4.a], *Cheyer* discloses a "user interaction logic operable to" perform the functionality recited in limitation [22.a]. (*See* above at Sections VII.C, IX.A.3.i, IX.A.11.iii.) Even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement logic was routine and commonplace at the time of the

alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

#### ii) [22.b] "query refining logic operable to refine the navigation query based upon the additional input; wherein the navigation logic users the refined navigation query to select a portion of the electronic data source"

163. I have been asked to assume that claim 22 contains a typographical error and was intended to recite "uses" instead of "users." Under that assumption, it is my opinion that *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding limitation [4.b]. (*See* above at Section IX.A.3.ii.)

164. Additionally, because *Cheyer* discloses an application implemented in software (*see* above at Section IX.A.6.i for citations and analysis regarding the preamble of claim 10), and for the reasons discussed above for limitation [4.b], *Cheyer* discloses a "query refining logic operable to" perform the functionality recited in limitation [22.b]. (*See* above at Sections VII.C, IX.A.3.ii.) Even if *Cheyer* were found not to provide for such an implementation, the use of a processor to implement logic was routine and commonplace at the time of the

alleged invention, and would have been a predictable modification. (Ex. 1013, 4:11-62 (disclosing a "computer processor" and various functions performed by executing the processor), 6:29-30 ("A computer processor 12 . . . controls the overall operation of the system.").)

165. A person of ordinary skill would have been motivated the configure the *navigation logic* to use the refined navigation query in the manner recited in limitation [22.b], so that the navigation logic (discussed above in Section IX.A.11.v for limitation [19.d]) could select a portion of the electronic data source based on refined information provided by the user. A person of ordinary skill would have found this to be a predictable configuration that would have improved the operation of the navigation logic..

### 14. Claim 24

#### i) [24.a] "The system of claim 19, wherein the system operates with respect to multiple users."

166. *Cheyer* in combination with *Thrift* and *Shwartz* discloses this limitation for at least the same reasons as presented above regarding claim 6. (*See* above at Section IX.A.4.)

#### 15. Claims 26, 27

##### i) [26.a] "The system of claim 19, wherein the mobile information appliance is a portable computing device."

##### ii) [27.a] "The system of claim 26, wherein the portable computing device is a personal digital assistant."

167. *Cheyer* in combination with *Thrift* and *Shwartz* discloses these limitations for at least the same reasons as presented above regarding claims 8 and 9. (*See* above at Sections IX.A.5.i-ii.)

### B. *Cheyer*, *Shwartz*, *Thrift*, and *Dureau* Disclose or Suggest the Features of Claims 2, 11, and 20

168. I reviewed *Cheyer*, *Shwartz*, *Thrift*, and *Dureau*, and in my opinion, *Cheyer*, *Shwartz*, *Thrift*, and *Dureau* disclose or suggest all of the features of claims 2, 11, and 20 of the '718 patent. Below, I address each of these claims and their respective limitations.

#### 1. Claim 2

##### i) [2.a] "The method of claim 1, wherein the step of rendering the interpretation of the spoken request is performed [at the one or more network servers]."

169. I have been asked to assume that claim 2 of the '718 patent contains an error and that it requires the step of rendering the interpretation of the spoken request to be performed "at the one or more network servers" instead of "by the mobile information appliance" as printed in the '718 patent. Under that

assumption, it is my opinion that *Cheyer* combined with *Thrift*, *Shwartz*, and *Dureau*, discloses or suggests this limitation.

170. *Cheyer* discloses a "server machine which will manage . . . natural language processing and speech recognition for the application." (Ex. 1012, 6; *see also id.*, 4 ("The user interface must be light and fast enough to run on a handheld PDA while able to access applications and data that may require a more powerful machine."), 11 (disclosing a "speech recognition agent, running on a remote speech server.").) I have been asked to assume that the server at which the data source is located according to the preamble of claim 1 must also perform the step of rendering the interpretation of the spoken request as in claim 2. Under that assumption, it is my opinion that while *Cheyer* does not expressly disclose that the server at which the data source is located according to the preamble of claim 1 also performs speech recognition and natural language processing, a person of ordinary skill would have been motivated in view of *Dureau* to configure the combined *Cheyer-Shwartz-Thrift* process to implement such features.

171. *Dureau* "relates generally to interactive television systems" (Ex. 1016, 1:8-12) and discloses voice input to a set-top box coupled to a television. (*Id.*, Abstract ("[A] microphone is coupled to a set-top box. The microphone allows the user to input voice information which is digitized and conveyed to the server for

conversion into textual information."), 10:56-11:1 ("[T]he user can enter his information by voice. The user can use a microphone or a telephone handset to provide voice data to the system. The microphone may a special-purpose microphone for use with the interactive television system or it may be a telephone handset. A special-purpose microphone may be connected to the set-top box, or it may be built into a remote control for the system. A telephone handset may be connected to the set-top box, or it may be connected directly to the return path (i.e., telephone line.) The voice data is transmitted to the server, which uses voice recognition software to convert the voice data into textual data. The textual data is returned to the set-top box, where it can be displayed to the user."), FIG. 1 (reproduced below).)

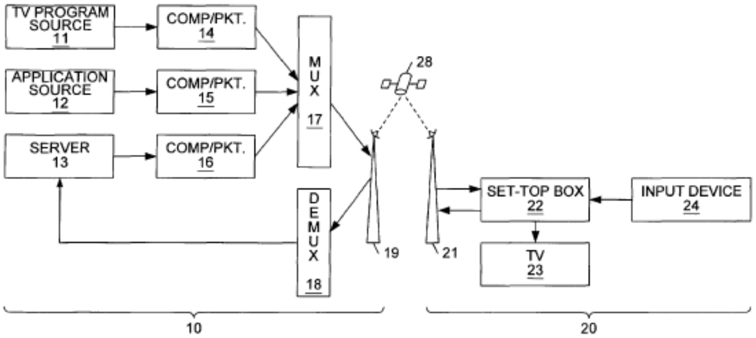


FIG. 1

(Ex. 1016, FIG. 1 (showing set-top box 22 connected to television 23.)

172. Because *Dureau*, like *Cheyer*, discloses that a user provides voice input that is processed by voice recognition software, a person of ordinary skill

would have had reason to consider the teachings of *Dureau* when implementing the *Cheyer-Shwartz-Thrift* process. Having looked to *Dureau*, a person of ordinary skill would have seen that *Dureau* discloses transmitting a user's speech input to a server, where it is interpreted, and further discloses performing applications relating to the speech input at the server. (Ex. 1016, Abstract ("[A] microphone is coupled to a set-top box. The microphone allows the user to input voice information which is digitized and conveyed to the server for conversion into textual information. The textual information is conveyed back to the set-top box and is input to an application executing on the set-top box."), 2:49-62 ("The invention comprises a system and method for enabling a user to provide non-textual information which is converted by the system to a textual form in which it can be used by the interactive application. The non-textual information is entered by the user at the set-top box of a receiving station and this information is transmitted to a server which may be located at a broadcast station. The server converts the information into textual data so that it can be used by the system. In one embodiment, the server transmits the textual data back to the receiving station, where it can be used by an application executing in the set-top box. In other embodiments, the textual data can be used at the server or transmitted to a part of the system other than the set-top box."), 3:39-44 ("The microphone is used to provide voice data, which is recorded and transmitted to a server equipped with a voice recognition application. The voice recognition application converts the voice data into textual data, which is then transmitted back to the application executing on the set-top box.”), 9:59-10:3 (“The system described above can be used with a number of different applications. For example, an interactive television service provider may wish to provide e-mail service to subscribers. The user can select the e-mail application furnished by the service provider and proceed to write the message which he or she wishes to send on the graphics tablet. In the message, the user writes the address of the intended recipient and the message to be sent to the recipient. The graphical data is transmitted to the server, which may segment the image data and then convert the data to text, or it may convert the entire image to text and then parse the text to determine the recipient’s address.”), 10:46-55 (“Another example of an application with which the system can be employed is electronic commerce service….”), 10:65-11:3 (“The voice data is transmitted to the server, which uses voice recognition software to convert the voice data into textual data. The textual data is returned to the set-top box, where it can be displayed to the user. The user can correct the text or confirm that the text has been accurately generated from the voice data.”).)

173. Based on *Dureau*'s disclosures regarding a server that is equipped with a voice recognition application and that performs applications using speech input, a person of ordinary skill would have been motivated to modify the combined *Cheyer-Thrift-Shwartz* process so that the server at which the data source is located as in the preamble of claim 1 also performs speech recognition and natural language processing. A person of ordinary skill would have known based on *Dureau* that such a configuration was possible, and he/she would have been motivated to implement the data source at the same server that performs speech recognition and natural language processing in order to achieve an efficient implementation. Such an implementation would have been a mere combination of known components and technologies, according to known methods, to achieve predictable results.

### 2. Claim 11

#### i) [11.a] "The computer program of claim 10, wherein the rendering of the interpretation of the spoken request is performed at the one or more network servers."

174. *Cheyer* in combination with *Thrift*, *Shwartz*, and *Dureau* discloses or suggests these limitations for at least the same reasons as presented above regarding claim 2. (*See* above at Section IX.B.1.)

#### 3. Claim 20

##### i) [20.a] "The system of claim 19, wherein the spoken language processing logic renders the interpretation of the spoken request at the one or more network servers."

175. *Cheyer* in combination with *Thrift*, *Shwartz*, and *Dureau* discloses or suggests these limitations for at least the same reasons as presented above regarding claim 2. (*See* above at Section IX.B.1.)

### C. *Cheyer*, *Shwartz*, *Thrift*, and *Johnson* Disclose or Suggest the Features of Claims 4, 13, and 22

176. I reviewed *Cheyer*, *Shwartz*, *Thrift*, and *Johnson*. I have been asked to consider a scenario in which claims 4, 13, and 22 require that the soliciting of additional input and refining of the navigation query recited in these claims must be for a different navigation query than the one recited in claim limitations [1.d], [10.d], and [19.d]. In such a scenario, it is my opinion that *Cheyer*, *Shwartz*, *Thrift*, and *Johnson* disclose or suggest all the features of claims 4, 13, and 22, as I discuss below. Such an approach was well within the skill of a person of ordinary skill and a mere design choice.

#### 4. Claim 4

##### i) [4.a] "The method of claim 1, further comprising the steps of soliciting additional input from the user, including user interaction in a modality different than the original request;"

177. As discussed above at paragraph 176, I have been asked to consider a scenario in which solicitation of additional input and refinement of the navigation query as in claim 4 must be for a navigation query that is different than the one recited in claim limitations [1.d], [10.d], and [19.d]. In such a scenario, it is my opinion that while *Cheyer*, *Thrift*, and *Shwartz* may not explicitly teach the steps of soliciting additional input from the user, including user interaction in a modality different than the original request, a person of ordinary skill in the art would have been motivated to modify the combined *Cheyer*-*Thrift*-*Shwartz* process (discussed above for claim 1) in view of *Johnson* to include such features. (*See* above at Section IX.A.1.)

178. *Johnson*, which is directed to "a multimodal natural language interface [that] interprets user requests," is in the same technical field as *Cheyer*. (Ex. 1014, Abstract.) A person of ordinary skill would have had reason to consider the teachings of *Johnson* for enhancing or augmenting the capabilities of the combined *Cheyer*-*Thrift*-*Shwartz* method, because *Cheyer*, *Thrift*, *Shwartz*, and *Johnson* are all directed to servicing user requests that are provided via an interface that includes natural language input.

179. *Johnson* discloses that in the example of a database query for "Joe Smith's telephone number," there could be "two Joe Smiths in the database," so

that "there is an ambiguity that must be clarified before a final response can be generated." (Ex. 1014, 5:7-18; *see also id.*, Abstract, 4:9-12, FIG. 4 (reproduced below).)

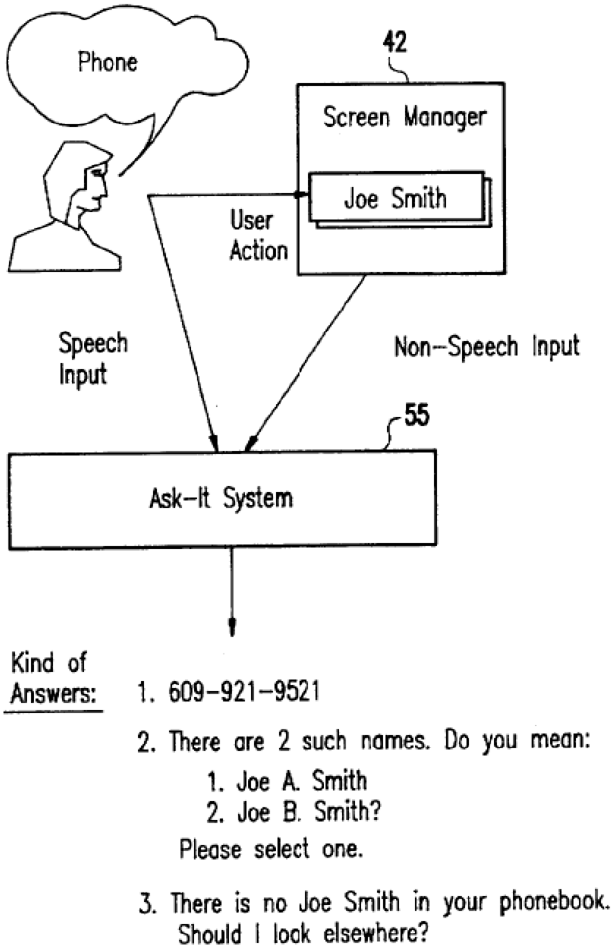


FIG.4

(*Id.*, FIG. 4.)

180. Thus, as shown in Figure 4, if there is an ambiguity, *Johnson*'s system asks the user to "select one" of the possibilities or indicate whether to look elsewhere. (Ex. 1014, FIG. 4.)

181. In view of *Johnson*'s disclosure of seeking clarification regarding an ambiguous situation in which two possible results are present, a person of ordinary skill would have been motivated to modify the combined *Cheyer*-*Thrift*-*Shwartz* process to clarify any ambiguity in a similar manner, and thus would have been motivated to solicit additional input from the user regarding such clarification, to provide the user with desired information in a user-friendly and convenient manner.

182. In view of *Johnson*'s disclosure of "provid[ing] a choice to the user … in a pop-up window, and request[ing] the user to select one of the choices" (Ex. 1014, 5:11-12), a person of ordinary skill would have been motivated to configure the combined process to include user interaction in a modality such as via a selection from a pop-up window in a graphical user interface without using voice input ("a modality different than the original request"), because such a skilled person would have recognized that providing the user with an ability to select a choice from a pop-up window by, for example, touching or clicking the choice

would have been a convenient, simple, and user-friendly implementation that would have enabled a wider range of input options for the user.

183. Indeed, *Cheyer* and *Johnson* in fact encourage such multimodal interaction, disclosing several examples in which the user provides non-spoken input. (*See, e.g.*, Ex. 1012, 1, 4, 5; Ex. 1014, Abstract, 2:21-22, 3:37-42, 3:44-46, 3:49-51.) Furthermore, input modalities other than speech input were well known long before the alleged invention of the ’718 patent.

184. In view of *Cheyer*’s and *Johnson*’s encouragement of multimodal input, a person of ordinary skill would have been motivated to modify the combined *Cheyer*-*Thrift*-*Shwartz* process to implement the features of limitation [4.a]. This modification would have been a mere combination of known components and technologies, according to known methods, to obtain predictable results.

#### ii) [4.b] “refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source”

185. As discussed above at paragraph 176, I have been asked to consider a scenario in which solicitation of additional input and refinement of the navigation query as in claim 4 must be for a navigation query that is different than the one recited in claim limitations [1.d], [10.d], and [19.d]. In such a scenario, it is my

opinion that while *Cheyer* and *Shwartz* may not explicitly teach refining the navigation query, based upon the additional input; wherein the at least one agent uses the refined navigation query to select a portion of the electronic data source, a person of ordinary skill in the art would have been motivated in view of *Johnson* to modify the combined *Cheyer*-*Shwartz* process to implement such features. As discussed above in Sections IX.A.1.iv-v and IX.A.4, *Cheyer* already discloses searching of an electronic data source based on a refined navigation query, just one that is refined based on ambiguity that is detected *before* the electronic data source is accessed. *Johnson* explicitly recognizes that ambiguities may be detected after an electronic data source is accessed, necessitating refinement of a navigation query and searching of an electronic data source after a first navigation query already searches the data source. For example, as discussed above in Section IX.C.1.i, *Johnson* discloses that in the example of a database query for "Joe Smith's telephone number," there could be "two Joe Smiths [found] in the database" after searching the database, so that "there is an ambiguity that must be clarified before a final response can be generated." (*See* above at Section IX.C.1.i; Ex. 1014, 5:7-18.) Thus, *Johnson* discloses requesting the user to select one of a plurality of choices or to specify whether a search should be conducted elsewhere. (Ex. 1014, FIG. 4.) As a result of the user's selection, *Johnson*'s system can find and present to the user the phone number that the user requested ("a portion of the electronic data source").

186. A person of ordinary skill would have been motivated to include in the combined *Cheyer-Thrift-Shwartz* process features such as those disclosed in *Johnson* regarding refining the navigation query after a database is initially searched based upon additional input and using the refined navigation query to select a portion of a database ("the electronic data source"), in order to enable the combined method and system to be able to handle situations where a user's request results in multiple, ambiguous hits or no hits at all. This would have been a simple modification for a person of ordinary skill to make, as it would have been merely a combination of known elements, according to known methods, to yield predictable results. (*See* above at Section IX.A.1.v for additional motivations to combine the references.) Indeed, a person of ordinary skill would have recognized that accessing and selecting a portion of an electronic data source with a refined navigation query would have involved substantially the same operations as compared to accessing and selecting a portion of an electronic data source with an original navigation query.

187. Indeed, a person of ordinary skill would have recognized the existence of two options for leveraging the user's clarification in *Johnson* to obtain the phone

number of the Joe Smith intended by the user: (a) access the database with a search query specifying "Joe Smith" and obtain an indication that there are two Joe Smiths in the database, without obtaining at that time the phone number for each Joe Smith (such that the phone number for the user-intended Joe Smith must later be retrieved from the database after the user's clarification); and (b) access the database with a search query specifying "Joe Smith" and obtain an indication that there are two Joe Smiths in the database, along with their respective phone numbers (such that upon the user's clarification, the user-intended phone number can simply be used without further accessing the database).

188. A person of ordinary skill would have recognized that configuring the combined *Cheyer-Thrift-Shwartz-Johnson* process to use the refined navigation query to select a portion of the electronic data source would have constituted a mere design choice among a finite number of known alternatives (e.g., the foregoing two options, which are not mutually exclusive, as a person of ordinary skill would have recognized that ambiguities could be resolved both before and after accessing the database), each having predictable outcomes (e.g., ultimately obtaining from the database the phone number of the user-intended Joe Smith).

189. I have been asked also to assume *Johnson* does not disclose the feature "to select a portion of the electronic data source." Under that assumption,

it is my opinion that a person of ordinary skill would have been motivated in view of *Shwartz* to implement that feature in the combined process for at least the same reasons presented above. (*See* above at Section IX.A.1.v.)

### 5. Claim 13

#### i) [13.a] "The computer program of claim 10, further comprising a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;"

#### ii) [13.b] "a code segment that refines the navigation query, based upon the additional input; and a code segment that uses the refined navigation query to select a portion of the electronic data source."

190. *Cheyer* in combination with *Thrift*, *Shwartz*, and *Johnson* discloses or suggests these limitations for at least the same reasons as presented above. (*See* above at Sections IX.A.6, IX.A.8.i-ii (citations and analysis regarding "code segment[s]"), IX.C.1.)

### 6. Claim 22

#### i) [22.a] "The system of claim 19, further comprising user interaction logic operable to solicit additional input from the user, including user interaction in a modality different than the original request; and"

#### ii) [22.b] "query refining logic operable to refine the navigation query based upon the additional input; wherein the navigation logic users the refined navigation query to select a portion of the electronic data source."

191. I have been asked to assume that claim 22 contains a typographical error and was intended to recite "uses" instead of "users." Under that assumption, it is my opinion that *Cheyer* in combination with *Thrift*, *Shwartz* and *Johnson* discloses or suggests these limitations for at least the same reasons as presented above. (*See* above at Sections IX.A.11, IX.A.13.i-ii (citations and analysis regarding "logic operable to"), IX.C.1.)

## D. *Cheyer*, *Shwartz*, *Thrift*, and *Simmers* Disclose or Suggest the Features of Claims 5, 7, 14, 16, 23, and 25

192. I reviewed *Cheyer*, *Shwartz*, *Thrift*, and *Simmers*, and, in my opinion, *Cheyer*, *Shwartz*, *Thrift*, and *Simmers* disclose or suggest all of the features of claims 5, 7, 14, 16, 23, and 25 of the '718 patent. Below, I address each of these claims and their respective limitations.

### 1. Claims 5, 7

#### i) [5.a] "The method of claim 1, wherein the data link includes a cellular telephone system."

#### ii) [7.a] "The method of claim 1, wherein the mobile information appliance is a wireless telephone."

193. While *Cheyer*, *Thrift*, and *Shwartz* do not expressly disclose a data link including a cellular telephone system, a person of ordinary skill would have been motivated in view of *Simmers* to configure the combined *Cheyer-Thrift-Shwartz* process to implement this feature.

194. As discussed above for claim 1, *Cheyer* discloses a data link between the user's mobile device and a remote server. (*See* above at Section IX.A.1.i.) However, because *Cheyer* is focused on other aspects of a process and system for obtaining information desired by a user, *Cheyer* does not provide details regarding the data link. *Cheyer* discloses that the mobile device can be a PDA (Ex. 1012, 4, 6), and a person of ordinary skill implementing *Cheyer*'s process would have recognized the desirability of incorporating cellular telephone functionality into a PDA. For example, *Simmers* discloses "dual-function information devices such as a cellular phone with PDA." (Ex. 1017, 1:47-48; *see also id.*, 1:12-15 ("'smart' cellular phones, which function both for telecommunications and for storing and retrieving information (e.g., a Personal Digital Assistant (information device))".)

195. A person of ordinary skill would have recognized that the mobile information appliance (e.g., a PDA with remote control functionality) in the combined *Cheyer-Thrift-Shwartz* process could have additionally been a wireless telephone, and would have been motivated to implement the device to be both a remote control device and a wireless telephone. For example, a person of ordinary skill would have recognized that the attributes of a PDA with added functionality of a remote control device and of a wireless telephone were not mutually exclusive, and that these were separate features that could have beneficially have

been co-implemented. Indeed, a person of ordinary skill would have been motivated to co-implement both of these features in order to provide a richer feature set for users. Such an implementation would have promoted efficiency, e.g., by using a single device to perform multiple features, and would have been consistent with the knowledge of a person of ordinary skill and the expectations of consumers regarding multi-function devices.

196. It was well known before the alleged invention of the ’718 patent that a cellular phone (e.g., as disclosed by *Simmers*) was used for communicating across a *cellular telephone system*. For example, *Haberman* discloses a cellular telecommunication system 1 (i.e., cellular telephone system or cellular network) including a mobile station 40 located in one cell and moving towards another cell. (Ex. 1018, 6:66-7:3 (“FIG. 1 shows a mobile station transitioning through a cellular telecommunication system according to the present invention including a CDMA portion of the cellular telecommunication system and an analog portion of the cellular telecommunication system.”), 7:12-14 (“FIG. 1 shows a cellular telecommunication system 1 according to the present invention”), 8:6-8 (“A mobile station 40 is located in a vehicle 45 that is currently in a digital cell 22 and moving towards an analog cell 21.”), FIG. 1 (reproduced below, and showing mobile station 40 in a cell of cellular telecommunication system 1).

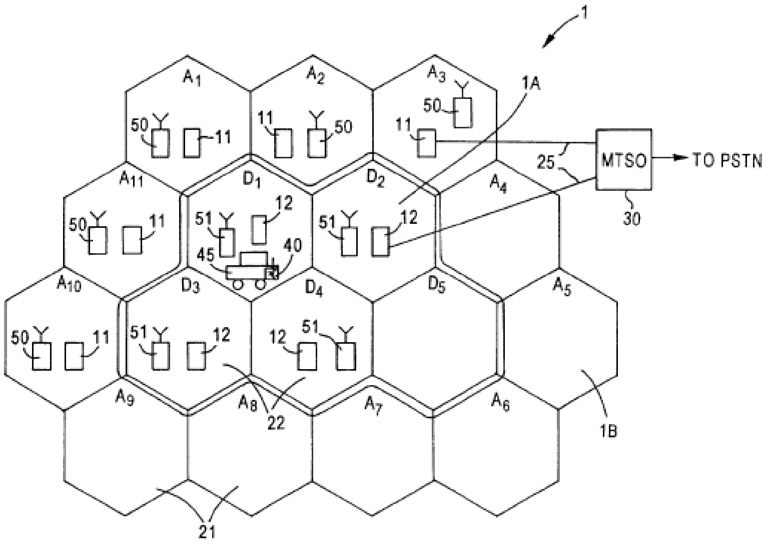


FIG. 1

(Ex. 1018, FIG. 1.)

197. A person of ordinary skill would have recognized the value of implementing a cellular telephone system (which a combination PDA/cellular phone as disclosed in *Simmers* would have used) to achieve a data link between *Cheyer*’s mobile device and remote data source. For example, a person of ordinary skill would have recognized that a cellular telephone system was a known system for communicating between a mobile device and a remote computer on the Web, and that *Cheyer* similarly discloses communications between a mobile device and a data source on the web (*see* above at Section IX.A.1.i), such that *Simmers*’s teachings regarding a combination PDA/cellular phone were relevant for implementing *Cheyer*’s communications. In view of *Simmers*’s teachings, a person of ordinary skill would have been motivated and been capable of modifying

the combined *Cheyer-Thrift-Shwartz* process so that *Cheyer*'s data link discussed above for the preamble of claim 1 includes a cellular telephone system.

198. This would have been a mere combination of known components and technologies (e.g., *Cheyer*'s disclosure of communication between a user's mobile device and a remote data source to provide the user with desired data from the data source, and *Simmers*'s disclosure of a cellular telephone system), according to known methods (e.g., a person of ordinary skill would have known how to implement a cellular telephone system to achieve *Cheyer*'s communication between a mobile device and a remote data source), to obtain predictable results (e.g., communication between two devices using a known networking technology).

199. Thus, a person of ordinary skill would have been motivated to implement *Cheyer*'s data link to include a cellular telephone system, as recited in claim 5.

200. A person of ordinary skill would further have been motivated in view of the foregoing references to configure *Cheyer*'s mobile device of the user ("mobile information appliance") to be a wireless telephone as recited in claim 7. For example, a person of ordinary skill would have known that a wireless telephone was typically used with a cellular telephone system in order to provide

portability and that a wireless telephone would have been a convenient device for a user to use and would have been familiar to the user.

201. Such a configuration would have been a mere combination of known components and technologies (e.g., a known cellular telephone system and a wireless telephone that was known to be used with such a cellular telephone system, *Cheyer*'s disclosure of a mobile device such as a PDA (Ex. 1012, 4, 6), and *Simmers*'s disclosure of a combination PDA-cell phone (Ex. 1017, 1:47-48)), according to known methods (e.g., a person of ordinary skill knew how to configure a device to be a wireless telephone), to achieve predictable results (e.g., providing a user with a wireless telephone).

202. A PDA's flexibility and expandability were well-known by the time of the alleged invention, and it was well-known that a PDA could operate as both a cellular phone (Ex. 1017, 1:47-48) and a remote control (Ex. 1033, 812). *Simmers* (discussed above) shows that it was known to implement a device that is both a cellular phone and a PDA (Ex. 1017, 1:47-48), and *Konstan* (discussed above for claims 8 and 9) shows that it was known to implement a device that is both a PDA and a remote control (Ex. 1033, 812). A person of ordinary skill would have similarly known how to implement a device that is both a remote control for a television and a wireless telephone. A person of ordinary skill would have

recognized the benefit of configuring a single device to be both a remote control for a television and a wireless telephone, e.g., so that the user could use the device to control his/her television when he/she was at home and could use the same device for cellular phone calls (e.g., when he/she was at home or in a car), thereby promoting convenience for the user in terms of reducing the number of devices that the user needed to use.

### 2. Claims 14, 16, 23, 25

#### i) [14.a] "The computer program of claim 10, wherein the data link includes a wireless telephone system."

#### ii) [16.a] "The computer program of claim 10, wherein the mobile information appliance is a wireless telephone."

#### iii) [23.a] "The system of claim 19, wherein the data link includes a cellular telephone system."

#### iv) [25.a] "The system of claim 19, wherein the mobile information appliance is a wireless telephone."

203. *Cheyer* in combination with *Thrift*, *Shwartz*, and *Simmers* discloses these limitations for at least the same reasons as presented above regarding claims 5 and 7. (*See* above at Sections IX.D.1.i-ii.) A person of ordinary skill would have been motivated to implement a *wireless* telephone system as recited in claim 14 for similar reasons as discussed above for claim 5 regarding implementing a *cellular*

telephone system, because a cellular telephone system was a type of wireless telephone system.

## X. CONCLUSION

204. I declare that all statements made herein of my knowledge are true, and that all statements made on information and belief are believed to be true, and that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: 12 JAN 18

By:____________________
Dr. Dan R. Olsen Jr.

# Dan R. Olsen Jr. - Vita

## education

Ph.D. in Computer and Information Science, University of Pennsylvania, 1981

M.S. in Computer Science, Brigham Young University, 1978

B. S. in Computer Science, Brigham Young University, 1976

## professional positions

Brigham Young University, Professor (1990-Present)

Carnegie Mellon University, Director, Human-Computer Interaction Institute (1996-1998)

Carnegie Mellon University, Professor, (1996-1998)

Brigham Young University, Department Chair of Computer Science, (1992-1996)

Brigham Young University, Associate Professor, (1986-1990)

Brigham Young University, Assistant Professor, (1984-1986)

Arizona State University, Assistant Professor, (1981-1984)

## awards

CHI Lifetime Research Award (2012)

Father of UIST Award (2007)

ACM Fellow (2007)

CHI Academy (2004)

Outstanding Paper Award (IUI 2003)

Karl G. Maeser Research and Creative Arts Award (2003)

Distinguished Paper Award (IUI 2002)

CHI Lifetime Service Award (2002)

## books

Building Interactive Systems, (2009)

Developing User Interfaces, (1998)

User Interface Management Systems: Models and Algorithms, (1992).

## papers

Dan Olsen, Benjamin Sellers, and Trent Boulter. "Enhancing interactive television news" TVX '14, ACM, pp 11-18 (June 2014)

Arthur, R., Olsen, D. R., "Window Brokers: Collaborative Display Space Control" ACM Transactions on Computer Human Interaction. Vol 19 (3), (Oct 2012)

Olsen, D. R., Bunn, D., Boulter, T., and Walz, R. "Interactive Television News" ACM Transactions on Multimedia Computing and Communications Applications. Vol 8 (2), (May 2012)

Arthur, R. and Olsen, D. R. "XICE Windowing Toolkit: Seamless Display Annexation" ACM Transactions on Computer-Human Interaction, Vol 18 (3), (July 2011)

Olsen, D. R. and Moon, B. "Video Summarization Based on User Interaction", EuroITV '11, ACM, pp 115-122, (June 2011)

Arthur, R. B. and Olsen, D. R. "Privacy-aware Shared UI Toolkit for Nomadic Environments" Software Practice and Experience, doi:10.1002/spe.1085 (May 2011)

Olsen, D. R., Partridge, B., and Lynn, S., "Time Warp Sports for Internet Television." ACM Transactions on Computer-Human Interaction, Vol 17 (4), (Dec 2010)

Lynn, S. G., Olsen, D. R., and Partridge, B. G., "Time Warp Football" European Conference on Interactive Television, ACM

(2009) (best paper award)

Olsen, D. R., and Harris, M. K., "Edge-Respecting Brushes" UIST 2008, ACM (2008)

Olsen, D. R. "Evaluating User Interface Systems Research," UIST 2007, ACM (2007)

Olsen, D. R., Clement, J., and Pace, A. "Spilling: Expanding Hand-held Interaction to Touch Table Displays", TableTop 2007, IEEE (2007)

Crandall, J., Goodrich, M., Nielsen, C., Olsen, D. R.:"Validating Human-Robot Interaction Schemes in Multitasking Environments", Systems, Man, and Cybernetics/Part A: Systems and Humans, Vol 35(4), IEEE, (July 2005)

Olsen, D. R., Taufer, T., Fails, J. A.: "ScreenCrayons: Annotating Anything", UIST '04, ACM, (2004)

Olsen, D. R., Wood, B.: "Fan-out: Measuring Human Control of Multiple Robots" CHI '04, ACM, (2004)

Olsen, D. R., Wood, B., Turner, J.: "Metrics for Human Driving of Multiple Robots" International Conference on Robots and Automation, ICRA '04, IEEE, (2004)

Goodrich, M., Olsen, D. R.: "Seven Principles of Efficient Human-Robot Interaction" International Conference on Systems, Man, and Cybernetics, IEEE, (2003)

Olsen, D. R., Goodrich, M.: "Metrics for Evaluating Human-Robot Interaction", PERMIS 2003

Fails, J. A., Olsen, D. R.: "A Design Tool for Camera-based Interaction", Human Factors in Computing Systems, CHI '03, ACM, (2003)

Fails, J. A., Olsen, D. R.: "Interactive Machine Learning", Intelligent User Interfaces, IUI '03, ACM (2003) (best paper award)

Olsen, D. R., Peachey, J. R.: "Query by Critique: Spoken Language Access to Large Lists", User Interface Software and

Technology, UIST '02, ACM(2002)

Fails, J. A., Olsen, D. R.: “Light Widgets: Interacting in Every-day Spaces”, Intelligent User Interfaces, IUI '02, ACM(2002)
(best paper award)

Rosenfeld, R., Olsen, D. R., Rudnicky, A. “Universal Speech Interfaces” interactions, ACM (2001)

Goodrich, M., Crandall, J. W., Palmer, T. J., Olsen, D. R.: “Experiments in Adjustable Autonomy,” Proceedings of the IJCAI01 workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents (2001)

Olsen, D. R., Nielsen, S. T., Parslow, D.: “Join and Capture: A Model for Nomadic Interaction,” UIST '01, ACM, (2001)

Olsen, D. R., Hudson, S., Tam, C. M., Conaty, G., Phelps, M., Heiner, J,: “Speech Interaction with Graphical User Interfaces”, Interact 2001, IOS Press, (2001)

Olsen, D. R., Nielsen, T,: “Laser pointer Interaction”, Human Factors in Computing Systems, CHI '01, ACM, (April 2001)

Olsen, D. R., Nielsen, T., Jefferies, S., Moyes, W., Fredrickson, P.: "Cross-modal Interaction in XWeb", UIST '00, ACM (2000)

Olsen, D. R., "Interacting in Chaos", Interactions, ACM (1999)

Olsen, D. R., Hudson, S. E., Verratti, T., Heiner, J. M., Phelps, M.: "Implementing Interface Attachments Based on Surface Representations", Human Factors in Computing Systems, CHI '99, ACM (1999)

Olsen, D. R., Hudson, S. E., Phelps, M., Heiner, J., Verratti, T.: “Ubiquitous Collaboration Via Surface Representations”, Computer Supported Cooperative Work, CSCW '98, ACM (1998)

Olsen, D. R., Boyarski, D., Verratti, T., Phelps, M., Lo, E.: “Generalized Pointing: Enabling Multiagent Interaction”, Human Factors in Computing Systems, CHI '98, ACM (1998)

Rodham, K.,Olsen, D.R.: “Nanites: An Approach to

Structure-Based Monitoring", ACM Transactions on Computer-Human Interaction, 4 (2), (1997)

Olsen, D. R., Deng, X.:"Inductive Groups" ACM Symposium on User Interface Software and Technology, UIST '96, (1996)

Nevill-Manning, C. G., Witten, I. H., Olsen, D. R.: "Compressing semi-structured text using hierarchical phrase identification", Proceedings of the Data Compression Conference, IEEE Press, Los Alamitos, CA (1996)

Olsen, D. R., Monk, A. F., Curry, M. B.: "Algorithms for automatic dialogue analysis using propositional production systems", Human Computer Interaction, 10, 39-78 (1995)

Olsen, D. R., Rodham, K.: "Interactive Net Services on the WWW" INTERACT '95. (1995)

Olsen, D. R., Rodham, K.: "Distributable Interactive Objects" INTERACT '95. (1995)

Olsen, D. R., Kohlert, D.: "Pictures as Input Data", Human Factors in Computing Systems, CHI '95, ACM (1995)

Olsen, D. R., Ahlstrom, B., Kohlert, D.: "Building Geometry-based Widgets by Example", Human Factors in Computing Systems, CHI '95, ACM (May 1995)

Olsen, D. R., Rodham, K.: "Smart Telepointers: Maintaining Telepointer Consistancy in the Presences of User Customization" ACM Transactions on Graphics (July 1994)

Olsen, D. R.:"Automatic Generation of Interactively Consistent Search Dialogs", Human Factors in Computing Systems, CHI '94, ACM (1994).

Olsen, D. R., Rodham, K., Kohlert, D.: "Implementing a Graphical Multi-user Interface Toolkit.", Software Practice and Experience (Sept 1993).

Foley, J. D., Hudson, S. E., Miller, J., Myers, B. A., Olsen, D. R. : "Research Directions for User Interface Software Tools", Behavior & Information Technology, (March-April 1993)

Olsen, D. R.: "Bookmarks: An enhanced scroll bar" ACM Transactions on Graphics, (July 1992)

Olsen, D. R.: "User Interface Architectures for an Information Age" HCI '92, York, United Kingdom, (Sept 1992)

Olsen, D. R., McNeill, T., Mitchell, D.: "Workspaces: an Architecture for editing collections of objects", Human Factors in Computing Systems, CHI '92, ACM (April 1992)

Becker, S., Barrett, W. A., Olsen, D. R.: "Interactive Measurement of Three-Dimensional Objects Using a Depth Buffer and Linear Probe", ACM Transactions on Graphics (1991)

Olsen, D. R., Burbidge, M.: "CTS - Complex Text Interaction System", Graphics Interface '90, (1990)

Olsen, D. R., Allen, K.: "Creating Interactive Techniques by Symbolically Solving Geometric Constraints", UIST '90, ACM (1990)

Olsen, D. R.: "Propositional Production Systems for Dialog Description", Human Factors in Computing Systems, CHI '90, ACM (1990)

Olsen, D. R., Tuck, R.: "Help by Guided Tasks - Utilizing UIMS Knowledge", Human Factors in Computing Systems, CHI '90, ACM (1990)

Olsen, D. R.: "A Programming Language Basis for User Interface Management", Human Factors in Computing Systems, CHI '89, ACM (1989)

Olsen, D. R., Burton, R. P.: "Structured Files for Interactive Graphics Programs", ISECON '88 Conference Proceedings. (1988)

Olsen, D. R., Halversen, B.: "Interface Usage Measurements in a User Interface Management System", ACM SIGGRAPH Symposium on User Interface Software, UIST '88, (Oct 1988)

Olsen, D. R.: "A Browse / Edit Model for User Interface Management" Graphics Interface '88. (June 1988)

Olsen, D. R., Dance, J.: "Macros by Example in a Graphical UIMS", IEEE Computer Graphics and Applications 6, 1 (Jan 1988)

Olsen, D. R., Burton, R. P.: "A Command-based User Interface Management System", 2nd International Conference on Computers and Applications (June 1988)

Olsen, D. R.: "Larger Issues in User Interface Management." Computer Graphics 21,2 (April 1987)

Olsen, D. R.: "MIKE: The Menu Interaction Kontrol Environment." ACM Transations on Graphics (Oct 1986)

Olsen, D. R., Nielsen, G.: "Direct Manipulation Techniques for 3D Objects Using 2D Locator Devices." 1986 Workshop on Interactive 3D Graphics, ACM (1986)

Olsen, D. R.: "Editing Templates: A User Interface Generation Tool." IEEE Computer Graphics and Applications 6, 11 (November 1986)

Olsen, D. R.: "An Editing Model for Generating Graphical User Interfaces." Proceedings of Graphics Interface '86. (May 1986)

Olsen, D. R., Dempsey, E., Rogge, R.: "Input-Output Linkage in a User Interface Management System.", Computer Graphics 19, 3 (July 1985)

Olsen, D. R., Cooper, C.: "Spatial Trees: A Fast Access Method for Unstructured Graphical Data.", Proceedings of Graphics Interface '85. (June 1985)

Olsen, D. R., Buxton, W., Ehrich, R., Kasik, D., Rhyne, J., Sibert, J.: "A Context for User Interface Management.", IEEE Computer Graphics and Applications 4,12 (December 1984)

Olsen, D. R.: "Push-down Automata for User Interface Management." ACM Transactions on Graphics 3,3 (July 1984)

Olsen, D. R.: "Presentational, Syntactic and Semantic Components of Interactive Dialogue Specification." Proceedings of the IFIP Workshop on User Interface Management. (November 1983)

Olsen, D. R., Dempsey, E.: "SYNGRAPH: An Automatic Interaction Generator.", Computer Graphics 17, 3 (July 1983)

Olsen, D. R., Dempsey, E.: "Syntax Directed Graphical Interaction.", Proceedings of SIGPLAN '83. (June 1983)

Olsen, D. R.: "Automatic Generation of Interactive Systems." Computer Graphics 17, 1 (January 1983)

Olsen, D. R.: "An Expression Model for Graphical Command Languages." Conference Proceedings National Computer Graphics Association, Vol II. (June 1982)

Olsen, D. R., Carter, S., Rockwood, A.: "A Program for Computing K-determinacy of a Taylor's Series.", Catastrophe Theory and its Applications. Appendix 1, pages 431-441. Pitman Publishing Limited, London, (1978)

## patents

Distributing Multiple Client Windows Using a Display Server - US 8671360 - issued March 11, 2014

Presentation Scratch Spaces - US 8356256 B2 - issued Jan 15, 2013

Window Broker Architecture - US 8171423 B2 - Issued May 1, 2012

## service positions

Adjunct Chair of Communities: ACM-SIGCHI - 2013-present

Vice President of Publications: ACM-SIGCHI - 2006-2013

Board Member: Utah Science Technology and Research (USTAR) governing board - 2006-2012

Director: Human-Computer Interaction Institute - Carnegie Mellon University - 1996-1998

Editor-in-Chief: ACM Transactions on Computer-Human Interaction - 1994-1997

Chair: Computer Science Department - Brigham Young University - 1992-1996

Chair: Faculty Advisory Council - Brigham Young University - 1988-1989

| ISSUE CLASSIFICATION | |
|---|---|
| Class | Subclass |
| 709 | 218 |

**PATENT NUMBER**

6757718

6757718

## U.S. UTILITY Patent Application

O.I.P.E.

SCANNED WJK Q.A. A.G.

PATENT DATE

JUN 29 2004

| APPLICATION NO. | CONT/PRIOR | CLASS | SUBCLASS | ART UNIT | EXAMINER |
|---|---|---|---|---|---|
| 09/608872 | D | ~~704~~ 709 | 218 | ~~2641~~ 5<br>2141 | ~~BACKER~~<br>Jean |

APPLICANTS
Christine Halverson
Luc Julia
Dimitris Voutsas
Adam Cheyer

TITLE
Mobile navigation of network-based electronic information using spoken input

PTO-2040
12/99

## ISSUING CLASSIFICATION

| ORIGINAL | | CROSS REFERENCE(S) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CLASS | SUBCLASS | CLASS | SUBCLASS (ONE SUBCLASS PER BLOCK) | | | | | |
| 709 | 218 | 709 | 202 | 217 | 219 | 227 | | |
| INTERNATIONAL CLASSIFICATION | | 704 | 257 | | | | | |
| G06F 15/16 | | | | | | | | |

☐ Continued on Issue Slip Inside File Jacket

☐ **TERMINAL DISCLAIMER**

☐ The term of this patent subsequent to __________ (date) has been disclaimed.

☐ The term of this patent shall not extend beyond the expiration date of U.S Patent. No. __________

☐ The terminal ____ months of this patent have been disclaimed.

**DRAWINGS**

| Sheets Drwg. | Figs. Drwg. | Print Fig. |
|---|---|---|
| 7 | 7 | 1A |

(Assistant Examiner) (Date)

Frantz B Jean 3/7/03
(Primary Examiner) (Date)

L. Johnson 3-13-03
(Legal Instruments Examiner) (Date)

**CLAIMS ALLOWED**

| Total Claims | Print Claim for O.G. |
|---|---|
| 27 | 1 |

**NOTICE OF ALLOWANCE MAILED**

3-11-03

**ISSUE FEE**

| Amount Due | Date Paid |
|---|---|
| $1650.00 | 5/6/03 |

**ISSUE BATCH NUMBER**

**WARNING:**
The information disclosed herein may be restricted. Unauthorized disclosure may be prohibited by the United States Code Title 35, Sections 122, 181 and 368. Possession outside the U.S. Patent & Trademark Office is restricted to authorized employees and contractors only.

Form PTO-436A
(Rev. 6/99)

FILED WITH: ☐ DISK (CRF) ☐ FICHE ☐ CD-ROM
(Attached in pocket on right inside flap)

ISSUE FEE IN FILE

(FACE)

4th

**PATENT APPLICATION**

09608372

846 U.S. PTO
09/608872

INITIALS 7/10/00 12

# CONTENTS

| | | Date Received (Incl. C. of M.) or Date Mailed | | | Date Received (Incl. C. of M.) or Date Mailed |
|---|---|---|---|---|---|
| 1. | Application ________ papers. | 8/31/00 | 42. | | |
| 2. | [illegible] | | 43. | | |
| 3. | Fee Surchage | 11-2-00 | 44. | | |
| 4. | Prel. Amdt A | 6-30-00 | 45. | | |
| 5. | Prel. Amdt B | 6-30-00 | 46. | | |
| 6. | 4-20-01 Rejection 3months | 4/24/01 | 47. | | |
| 7. | IDS | 4-30-01 | 48. | | |
| 8. | Associate P/A | 5-14-01 | 49. | | |
| 9. | Change of address | 5-14-01 | 50. | | |
| 10. | Ext. of time ② | 9-21-01 | 51. | | |
| 11. | Revocation & Power of Attorney | 9-21-01 | 52. | | |
| 12. | Req. for reconsideration | 9-21-01 | 53. | | |
| 13. | notice of acceptance | 10-2-01 | 54. | | |
| 14. | 10-4-01 Final Rejection 3months | 10/10/01 | 55. | | |
| 15. | Interview Summary | 1-16-02 | 56. | | |
| 16. | Amdt C | 1-10-02 | 57. | | |
| 17. | 1-28-02 Advisory Action | 1-28-02 | 58. | | |
| 18. | Ext of time/month / Request for RCE | 2-~~8~~-02 | 59. | | |
| 19. | 2-19-02 Rejection 3months | 2-19-02 | 60. | | |
| 20. | Interview Summary | | 61. | | |
| 21. | Ext of time (2mos) | 7/18/02 | 62. | | |
| 22. | Amdt D | 7/18/02 | 63. | | |
| 23. | IDS | 7/24/02 | 64. | | |
| 24. | Rejection 3months | 10-4-02 | 65. | | |
| 25. | Response | 1-6-03 | 66. | | |
| 26. | Interview Summary | 01/08/03 | 67. | | |
| 27. | 3-10-03 notice of allow. | 3/11/03 | 68. | | |
| 28. | Formal Drawings ( 7 shts) set 1 | 05-06-03 | 69. | | |
| 29. | | | 70. | | |
| 30. | | | 71. | | |
| 31. | | | 72. | | |
| 32. | | | 73. | | |
| 33. | | | 74. | | |
| 34. | | | 75. | | |
| 35. | | | 76. | | |
| 36. | | | 77. | | |
| 37. | | | 78. | | |
| 38. | | | 79. | | |
| 39. | | | 80. | | |
| | | | 81. | | |
| | | | 82. | | |

(LEFT OUTSIDE)

PATENT APPLICATION SERIAL NO. ____________________

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

```
07/12/2000 PALLEN   00000020 09608872

01 FC:201                        345.00 OP
02 FC:203                         63.00 OP
```

PTO-1556
(5/87)

*U.S. GPO: 1999-459-082/19144

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

**CONFIRMATION NO. 2382**

| SERIAL NUMBER | FILING DATE | CLASS | GROUP ART UNIT | ATTORNEY DOCKET NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000<br>RULE | 709 | 2155 | SRIlp037B |

**APPLICANTS**

Christine Halversen, San Jose, CA;
Luc Julia, Menlo Park, CA;
Dimitris Voutsas, Thessaloniki, GREECE;
Adam Cheyer, Palo Alto, CA;

** **CONTINUING DATA** *************************** ok FO

THIS APPLICATION IS A CON OF 09/524,095 03/13/2000
WHICH IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
AND SAID 09/524,095 03/13/2000
CLAIMS BENEFIT OF 60/124,720 03/17/1999
AND CLAIMS BENEFIT OF 60/124,719 03/17/1999

** **FOREIGN APPLICATIONS** *********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** SMALL ENTITY **
** 08/31/2000

Foreign Priority claimed ☐ yes ☒ no
35 USC 119 (a-d) conditions met ☐ yes ☑ no ☐ Met after Allowance
Verified and Acknowledged ________ Examiner's Signature ________ Initials

| STATE OR COUNTRY | SHEETS DRAWING | TOTAL CLAIMS | INDEPENDENT CLAIMS |
|---|---|---|---|
| CA | 7 | 27 | 3 |

**ADDRESS**

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

**TITLE**

mobile navigation of network-based electronic information using spoken input

**FILING FEE RECEIVED**
473

FEES: Authority has been given in Paper No. ________ to charge/credit DEPOSIT ACCOUNT No. ________ for following:

- ☐ All Fees
- ☐ 1.16 Fees ( Filing )
- ☐ 1.17 Fees ( Processing Ext. of time )
- ☐ 1.18 Fees ( Issue )
- ☐ Other ______________
- ☐ Credit

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

| SERIAL NUMBER | FILING DATE | CLASS | GROUP ART UNIT | ATTORNEY DOCKET NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000<br>RULE – | 704 | 2741 | SRIlp037B |

**APPLICANTS**

Christine Halversen, San Jose, CA ;
Luc Julia, Menlo Park`, CA ;
Dimitris Voutsas, Thessaloniki, GREECE;
Adam Cheyer, Palo Alto, CA ;

✓

** CONTINUING DATA ************************* yes FJ

THIS APPLICATION IS A CON OF 09/524,095 03/13/2000
WHICH IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,719 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,720 03/17/1999

** FOREIGN APPLICATIONS ******************** None FJ

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** 08/31/2000

** SMALL ENTITY **

| | STATE OR COUNTRY | SHEETS DRAWING | TOTAL CLAIMS | INDEPENDENT CLAIMS |
|---|---|---|---|---|
| Foreign Priority claimed ☐ yes ☒ no<br>35 USC 119 (a-d) conditions met ☐ yes ☒ no ☐ Met after Allowance<br>Verified and Acknowledged FJ Examiner's Signature FJ Initials | CA | 7 | 27 | 3 |

**ADDRESS**

24277

**TITLE**

Mobile navigation of network-based electronic information using spoken input

| | | |
|---|---|---|
| **FILING FEE RECEIVED**<br>473 | FEES: Authority has been given in Paper No. ________ to charge/credit DEPOSIT ACCOUNT No. ________ for following: | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other ____________<br>☐ Credit |

JC846 U.S. PTO
09/608872
06/30/00

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this paper and the documents and/or fees referred to as attached therein are being deposited with the United States Postal Service on June 30, 2000 in an envelope as "Express Mail Post Office to Addressee" service under 37 CFR §1.10, Mailing Label Number **EK858788212US**, addressed to the Assistant Commissioner for Patents, Washington, DC 20231.

Kevin J. Zilka

Attorney Docket No.: SRI1P037B

First Named Inventor:

HALVERSEN, Christine

## UTILITY PATENT APPLICATION TRANSMITTAL (37 CFR. § 1.53(b))

(Continuation, Divisional or Continuation-in-part application)

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

☒ Duplicate for fee processing

Sir: This is a request for filing a patent application under 37 CFR. § 1.53(b) in the name of inventors: Christine Halversen, Luc Julia, Dimitris Voutsas, Adam Cheyer

For: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

This application is a ☒ Continuation ☐ Divisional ☐ Continuation-in-part

of prior Application No.: **09/524,095**, from which priority under 35 U.S.C. §120 is claimed.

Application Elements:

☒ 33 Pages of Specification, Claims and Abstract
☒ 07 Sheets of Drawings
☐ Declaration
    ☐ Newly executed (original or copy)
    ☐ Copy from a prior application (37 CFR 1.63(d) for a continuation or divisional). The entire disclosure of the prior application from which a copy of the declaration is herein supplied is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
        ☐ Deletion of inventors Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).

Accompanying Application Parts:

☐ Assignment and Assignment Recordation Cover Sheet (recording fee of $40.00 enclosed)
☐ Power of Attorney
☐ 37 CFR 3.73(b) Statement by Assignee

☐ Information Disclosure Statement with Form PTO-1449 ☐ Copies of IDS Citations

☒ Preliminary Amendment

☒ Return Receipt Postcard

☐ Small Entity Statement(s) ☒ Statement filed in prior application. Status still proper and desired.

☐ Other:

Claim For Foreign Priority

☐ Priority of ________ Application No. _________ filed on __________ is claimed under 35 U.S.C. § 119.

☐ The certified copy has been filed in prior application U.S. Application No. _______

☐ The certified copy will follow.

Extension of Time for Prior Pending Application

☐ A Petition for Extension of Time is being concurrently filed in the prior pending application. A copy of the Petition for Extension of Time is attached.

Amendments

☐ Amend the specification by inserting before the first line the sentence: "This is a

☐ Continuation ☐ Continuation-in-part ☐ Divisional

application of copending prior

☐ Application No._________ filed on _____________,

☐ International Application __________ filed on _________ which designated the United States,

the disclosure of which is incorporated herein by reference."

☒ Cancel in this application original claims 2-55 of the prior application before calculating the filing fee. (*At least one original independent claim must be retained.*)

Fee Calculation (37 CFR § 1.16)

| | (Col. 1) NO. FILED | (Col. 2) NO. EXTRA | SMALL ENTITY RATE | FEE | OR | LARGE ENTITY RATE | FEE |
|---|---|---|---|---|---|---|---|
| BASIC FEE | | | $345 $ | 345 | OR | $690 $ | |
| TOTAL CLAIMS | 27 -20 = | 7 | x09 = $ | 63 | OR | x18 = $ | |
| INDEP CLAIMS | 3 -03 = | 0 | x39 = $ | | OR | x78 = $ | |
| [ ] Multiple Dependent Claim Presented | | | $130 = $ | | OR | $260 = $ | |
| * If the difference in Col. 1 is less than zero, enter "0" in Col. 2. | | | Total $ | 408 | OR | Total $ | |

☒ Check No. 137 in the amount of $ 408.00 is enclosed.

☒ The Commissioner is authorized to charge any fees beyond the amount enclosed which may be required, or to credit any overpayment, to Deposit Account No. 50-1351 (Order No. SRI1P037B).

General Authorization for Petition for Extension of Time (37 CFR §1.136)

☒ Applicants hereby make and generally authorize any Petitions for Extensions of Time as may be needed for any subsequent filings. The Commissioner is also authorized to charge any extension fees under 37 CFR §1.17 as may be needed to Deposit Account No. 50-1351 (Order No. SRI1P037B).

☒ Please send correspondence to the following address:

**Kevin J. Zilka**
**P.O. BOX 721030**
**San Jose, California 95172-1030**

**Direct Telephone Calls To:** **Kevin J. Zilka at telephone number (408) 505-5100**

Date: June 30, 2000

**Kevin J. Zilka**
Registration No. 41,429

(Revised 12/97, Pat App Trans 53(b) ContDivCIP)

# NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

## BACKGROUND OF THE INVENTION

rns B2

~~This is~~ 185 a Continuation In Part of co-pending U.S. Patent Application No. 09/225,198, filed January 5, 1999, Provisional U.S. Patent Application No. 60/124,718, filed March 17, 1999, Provisional U.S. Patent Application No. 60/124,720, filed March 17, 1999, and Provisional U.S. Patent Application No. 60/124,719, filed March 17, 1999, from which applications priority is claimed and these application are incorporated herein by reference.

The present invention relates generally to the navigation of electronic data by means of spoken natural language requests, and to feedback mechanisms and methods for resolving the errors and ambiguities that may be associated with such requests.

As global electronic connectivity continues to grow, and the universe of electronic data potentially available to users continues to expand, there is a growing need for information navigation technology that allows relatively naïve users to navigate and access desired data by means of natural language input. In many of the most important markets -- including the home entertainment arena, as well as mobile computing -- spoken natural language input is highly desirable, if not ideal. As just one example, the proliferation of high-bandwidth communications infrastructure for the home entertainment market (cable, satellite, broadband) enables delivery of movies-on-demand and other interactive multimedia content to the consumer's home television set. For users to take full advantage of this content stream ultimately requires interactive navigation of content databases in a manner that is too complex for user-friendly selection by means of a traditional remote-control clicker. Allowing spoken natural language requests as the input modality for rapidly searching and accessing desired content is an important objective for a successful consumer entertainment product in a context offering a dizzying range of database content choices. As further examples, this same need to drive navigation of (and transaction with) relatively complex data warehouses using spoken natural language requests applies equally to surfing the Internet/Web or other networks for general information, multimedia content, or e-commerce transactions.

2

In general, the existing navigational systems for browsing electronic databases and data warehouses (search engines, menus, etc.), have been designed without navigation via spoken natural language as a specific goal. So today's world is full of existing electronic data navigation systems that do not assume browsing via natural spoken commands, but rather assume text and mouse-click inputs (or in the case of TV remote controls, even less). Simply recognizing voice commands within an extremely limited vocabulary and grammar -- the spoken equivalent of button/click input (e.g., speaking "channel 5" selects TV channel 5) -- is really not sufficient by itself to satisfy the objectives described above. In order to deliver a true "win" for users, the voice-driven front-end must accept spoken natural language input in a manner that is intuitive to users. For example, the front-end should not require learning a highly specialized command language or format. More fundamentally, the front-end must allow users to speak directly in terms of what the user ultimately wants -- e.g., "I'd like to see a Western film directed by Clint Eastwood" -- as opposed to speaking in terms of arbitrary navigation structures (e.g., hierarchical layers of menus, commands, etc.) that are essentially artifacts reflecting constraints of the pre-existing text/click navigation system. At the same time, the front-end must recognize and accommodate the reality that a stream of naïve spoken natural language input will, over time, typically present a variety of errors and/or ambiguities: e.g., garbled/unrecognized words (did the user say "Eastwood" or "Easter"?) and under-constrained requests ("Show me the Clint Eastwood movie"). An approach is needed for handling and resolving such errors and ambiguities in a rapid, user-friendly, non-frustrating manner.

What is needed is a methodology and apparatus for rapidly constructing a voice-driven front-end atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the step-by-step browsing architecture of the existing navigation system, and wherein any errors or ambiguities in user input are rapidly and conveniently resolved. The solution to this need should be compatible with the constraints of a multi-user, distributed environment such as the Internet/Web or a proprietary high-bandwidth content delivery network; a solution contemplating one-at-a-time user interactions at a single location is insufficient, for example.

3

## SUMMARY OF THE INVENTION

Sub B3

The present invention addresses the above needs by providing a system, method, and article of manufacture for navigating network-based electronic data sources in response to spoken NL input requests. When a spoken natural language input request is received from a user, it is interpreted, such as by using a speech recognition engine to extract speech data from acoustic voice signals, and using a natural language parser to linguistically parse the speech data. The interpretation of the spoken natural language request can be performed on a computing device locally with the user or remotely from the user. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is then transmitted to a client device of the user. If the network data source is a database, the navigation query is constructed in the format of a database query language.

Typically, errors or ambiguities emerge in the interpretation of the spoken NL request, such that the system cannot instantiate a complete, valid navigational template. This is to be expected occasionally, and one preferred aspect of the invention is the ability to handle such errors and ambiguities in relatively graceful and user-friendly manner. Instead of simply rejecting such input and defaulting to traditional input modes or simply asking the user to try again, a preferred embodiment of the present invention seeks to converge rapidly toward instantiation of a valid navigational template by soliciting additional clarification from the user as necessary, either before or after a navigation of the data source, via multimodal input, i.e., by means of menu selection or other input modalities including and in addition to spoken natural language. This clarifying, multi-modal dialogue takes advantage of whatever partial navigational information has been gleaned from the initial interpretation of the user's spoken NL request. This clarification process continues until the system converges toward an adequately instantiated navigational template, which is in turn used to navigate the network-based data and retrieve the user's desired information. The retrieved information is transmitted across the network and presented to the user on a suitable client display device.

4

In a further aspect of the present invention, the construction of the navigation query includes extracting an input template for an online scripted interface to the data source and using the input template to construct the navigation query. The extraction of the input template can include dynamically scraping the online scripted interface.

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention, together with further advantages thereof, may best be understood by reference to the following description taken in conjunction with the accompanying drawings in which:

Figure 1a illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with server-side processing of requests;

Figure 1b illustrates another system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with client-side processing of requests;

Figure 2 illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention for a mobile computing scenario;

Figure 3 illustrates the functional logic components of a request processing module in accordance with an embodiment of the present invention;

Figure 4 illustrates a process utilizing spoken natural language for navigating an electronic database in accordance with one embodiment of the present invention;

Figure 5 illustrates a process for constructing a navigational query for accessing an online data source via an interactive, scripted (e.g., CGI) form; and

Figure 6 illustrates an embodiment of the present invention utilizing a community of distributed, collaborating electronic agents.

6

## DETAILED DESCRIPTION OF THE INVENTION

### 1. System Architecture

a. Server-End Processing of Spoken Input

Figure 1a is an illustration of a data navigation system driven by spoken natural language input, in accordance with one embodiment of the present invention. As shown, a user's voice input data is captured by a voice input device 102, such as a microphone. Preferably voice input device 102 includes a button or the like that can be pressed or held-down to activate a listening mode, so that the system need not continually pay attention to, or be confused by, irrelevant background noise. In one preferred embodiment well-suited for the home entertainment setting, voice input device 102 is a portable remote control device with an integrated microphone, and the voice data is transmitted from device 102 preferably via infrared (or other wireless) link to communications box 104 (e.g., a set-top box or a similar communications device that is capable of retransmitting the raw voice data and/or processing the voice data) local to the user's environment and coupled to communications network 106. The voice data is then transmitted across network 106 to a remote server or servers 108. The voice data may preferably be transmitted in compressed digitized form, or alternatively --particularly where bandwidth constraints are significant-- in analog format (e.g., via frequency modulated transmission), in the latter case being digitized upon arrival at remote server 108.

At remote server 108, the voice data is processed by request processing logic 300 in order to understand the user's request and construct an appropriate query or request for navigation of remote data source 110, in accordance with the interpretation process exemplified in Figure 4 and Figure 5 and discussed in greater detail below. For purposes of executing this process, request processing logic 300 comprises functional modules including speech recognition engine 310, natural language (NL) parser 320, query construction logic 330, and query refinement logic 340, as shown in Figure 3. Data source 110 may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably resides on a central server or servers -- which may or may not be the same as server 108, depending on the storage and bandwidth needs of the application and the resources available to the practitioner. Data source 110 may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source 110 are navigated -- i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user -- using the processes of Figures 4 and 5 as described in greater detail below.

Once the desired information has been retrieved from data source 110, it is electronically transmitted via network 106 to the user for viewing on client display device 112. In a preferred embodiment well-suited for the home entertainment setting, display device 112 is a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device 112 is coupled to or integrated with a communications box (which is preferably the same as communications box 104, but may also be a separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network 106.

Network 106 is a two-way electronic communications network and may be embodied in electronic communication infrastructure including coaxial (cable television) lines, DSL, fiber-optic cable, traditional copper wire (twisted pair), or any other type of hardwired connection. Network 106 may also include a wireless connection such as a satellite-based connection, cellular connection, or other type of wireless connection. Network 106 may be part of the Internet and may support TCP/IP communications, or may be embodied in a proprietary network, or in any other electronic communications network infrastructure, whether packet-switched or connection-oriented. A design consideration is that network 106 preferably provide suitable bandwidth depending upon the nature of the content anticipated for the desired application.

### b. Client-End Processing of Spoken Input

Figure 1b is an illustration of a data navigation system driven by spoken natural language input, in accordance with a second embodiment of the present invention. Again, a user's voice input data is captured by a voice input device 102, such as a microphone. In the embodiment shown in Figure 1b, the voice data is transmitted from device 202 to requests processing logic 300, hosted on a local speech processor, for processing and interpretation. In the preferred embodiment illustrated in Figure 1b, the local speech processor is conveniently integrated as part of communications box 104, although implementation in a physically separate (but communicatively coupled) unit is also possible as will be readily apparent to those of skill in the art. The voice data is processed by the components of request processing logic 300 in order to understand the user's request and construct an appropriate query or request for navigation of remote data source 110, in accordance with the interpretation process exemplified in Figures 4 and 5 as discussed in greater detail below.

The resulting navigational query is then transmitted electronically across network 106 to data source 110, which preferably resides on a central server or servers 108. As in Figure 1a, data source 110 may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source 110 are then navigated -- i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user -- preferably using the process of Figures 4 and 5 as described in greater detail below. Once the desired information has been retrieved from data source 110, it is electronically transmitted via network 106 to the user for viewing on client display device 112.

In one embodiment in accordance with Figure 1b and well-suited for the home entertainment setting, voice input device 102 is a portable remote control device with an integrated microphone, and the voice data is transmitted from device 102 preferably via infrared (or other wireless) link to the local speech processor. The local speech processor is coupled to communications network 106, and also preferably to client display device 112 (especially for purposes of query refinement transmissions, as discussed below in connection with Figure 4, step 412), and preferably may be integrated within or coupled to communications box 104. In addition, especially for purposes of a home entertainment application, display device 112 is preferably a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device 112 is coupled to a communications box (which is preferably the same as communications box 104, but may also be a physically separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network 106.

Design considerations favoring server-side processing and interpretation of spoken input requests, as exemplified in Figure 1a, include minimizing the need to distribute costly computational hardware and software to all client users in order to perform speech and language processing. Design considerations favoring client-side processing, as exemplified in Figure 1b, include minimizing the quantity of data sent upstream across the network from each client, as the speech recognition is performed before transmission across the network and only the query data and/or request needs to be sent, thus reducing the upstream bandwidth requirements.

c. Mobile Client Embodiment

A mobile computing embodiment of the present invention may be implemented by practitioners as a variation on the embodiments of either Figure 1a or Figure 1b. For example, as depicted in Figure 2, a mobile variation in accordance with the server-side processing architecture illustrated in Figure 1a may be implemented by replacing voice input device 102, communications box 104, and client display device 112, with an integrated, mobile, information appliance 202 such as a cellular telephone or wireless personal digital assistant (wireless PDA). Mobile information appliance 202 essentially performs the functions of the replaced components. Thus, mobile information appliance 202 receives spoken natural language input requests from the user in the form of voice data, and transmits that data (preferably via wireless data receiving station 204) across communications network 206 for server-side interpretation of the request, in similar fashion as described above in connection with Figure 1. Navigation of data source 210 and retrieval of desired information likewise proceeds in an analogous manner as described above. Display information transmitted electronically back to the user across network 206 is displayed for the user on the display of information appliance 202, and audio information is output through the appliance's speakers.

Practitioners will further appreciate, in light of the above teachings, that if mobile information appliance 202 is equipped with sufficient computational processing power, then a mobile variation of the client-side architecture exemplified in Figure 2 may similarly be implemented. In that case, the modules corresponding to request processing logic 300 would be embodied locally in the computational resources of mobile information appliance 202, and the logical flow of data would otherwise follow in a manner analogous to that previously described in connection with Figure 1b.

As illustrated in Figure 2, multiple users, each having their own client input device, may issue requests, simultaneously or otherwise, for navigation of data source 210. This is equally true (though not explicitly drawn) for the embodiments depicted in Figures 1a and 1b. Data source 210 (or 100), being a network accessible information resource, has typically already been constructed to support access requests from simultaneous multiple network users, as known by practitioners of ordinary skill in the art. In the case of server-side speech processing, as exemplified in Figures 1a and 2, the interpretation logic and error correction logic modules are also preferably designed and implemented to support queuing and multi-tasking of requests from multiple simultaneous network users, as will be appreciated by those of skill in the art.

It will be apparent to those skilled in the art that additional implementations, permutations and combinations of the embodiments set forth in Figures 1a, 1b, and 2 may be created without straying from the scope and spirit of the present invention. For example, practitioners will understand, in light of the above teachings and design considerations, that it is possible to divide and allocate the functional components of request processing logic 300 between client and server. For example, speech recognition -- in entirety, or perhaps just early stages such as feature extraction -- might be performed locally on the client end, perhaps to reduce bandwidth requirements, while natural language parsing and other necessary processing might be performed upstream on the server end, so that more extensive computational power need not be distributed locally to each client. In that case, corresponding portions of request processing logic 300, such as speech recognition engine 310 or portions

11

thereof, would reside locally at the client as in Figure 1b, while other component modules would be hosted at the server end as in Figures 1a and 2.

Further, practitioners may choose to implement the each of the various embodiments described above on any number of different hardware and software computing platforms and environments and various combinations thereof, including, by way of just a few examples: a general-purpose hardware microprocessor such as the Intel Pentium series; operating system software such as Microsoft Windows/CE, Palm OS, or Apple Mac OS (particularly for client devices and client-side processing), or Unix, Linux, or Windows/NT (the latter three particularly for network data servers and server-side processing), and/or proprietary information access platforms such as Microsoft's WebTV or the Diva Systems video-on-demand system.

## 2. Processing Methodology

The present invention provides a spoken natural language interface for interrogation of remote electronic databases and retrieval of desired information. A preferred embodiment of the present invention utilizes the basic methodology outlined in the flow diagram of Figure 4 in order to provide this interface. This methodology will now be discussed.

### a. Interpreting Spoken Natural Language Requests

At step 402, the user's spoken request for information is initially received in the form of raw (acoustic) voice data by a suitable input device, as previously discussed in connection with Figures 1-2. At step 404 the voice data received from the user is interpreted in order to understand the user's request for information. Preferably this step includes performing speech recognition in order to extract words from the voice data, and further includes natural language parsing of those words in order to generate a structured linguistic representation of the user's request.

Speech recognition in step 404 is performed using speech recognition engine 310. A variety of commercial quality, speech recognition engines are readily available on the market, as practitioners will know. For example, Nuance Communications offers a suite of speech recognition engines, including Nuance 6, its current flagship product, and Nuance Express, a lower cost package for entry-level

12

applications. As one other example, IBM offers the ViaVoice speech recognition engine, including a low-cost shrink-wrapped version available through popular consumer distribution channels. Basically, a speech recognition engine processes acoustic voice data and attempts to generate a text stream of recognized words.

Typically, the speech recognition engine is provided with a vocabulary lexicon of likely words or phrases that the recognition engine can match against its analysis of acoustical signals, for purposes of a given application. Preferably, the lexicon is dynamically adjusted to reflect the current user context, as established by the preceding user inputs. For example, if a user is engaged in a dialogue with the system about movie selection, the recognition engine's vocabulary may preferably be adjusted to favor relevant words and phrases, such as a stored list of proper names for popular movie actors and directors, etc. Whereas if the current dialogue involves selection and viewing of a sports event, the engine's vocabulary might preferably be adjusted to favor a stored list of proper names for professional sports teams, etc. In addition, a speech recognition engine is provided with language models that help the engine predict the most likely interpretation of a given segment of acoustical voice data, in the current context of phonemes or words in which the segment appears. In addition, speech recognition engines often echo to the user, in more or less real-time, a transcription of the engine's best guess at what the user has said, giving the user an opportunity to confirm or reject.

In a further aspect of step 404, natural language interpreter (or parser) 320 linguistically parses and interprets the textual output of the speech recognition engine. In a preferred embodiment of the present invention, the natural-language interpreter attempts to determine both the meaning of spoken words (semantic processing) as well as the grammar of the statement (syntactic processing), such as the Gemini Natural Language Understanding System developed by SRI International. The Gemini system is described in detail in publications entitled "Gemini: A Natural Language System for Spoken-Language Understanding" and "Interleaving Syntax and Semantics in an Efficient Bottom-Up Parser," both of which are currently available online at http://www.ai.sri.com/natural-language/projects/arpa-sls/nat-lang.html. (Copies of those publications are also included in an information disclosure statement submitted herewith, and are incorporated herein by this reference). Briefly, Gemini applies a set of syntactic and semantic grammar rules to a word string using a bottom-up parser to generate a logical form, which is a structured representation of the context-independent meaning of the string. Gemini can be used with a variety of grammars, including general English grammar as well as application-specific grammars. The Gemini parser is based on "unification grammar," meaning that grammatical categories incorporate features that can be assigned values; so that when grammatical category expressions are matched in the course of parsing or semantic interpretation, the information contained in the features is combined, and if the feature values are incompatible the match fails.

It is possible for some applications to achieve a significant reduction in speech recognition error by using the natural-language processing system to re-score recognition hypotheses. For example, the grammars defined for a language parser like Gemini may be compiled into context-free grammar that, in turn, can be used directly as language models for speech recognition engines like the Nuance recognizer. Further details on this methodology are provided in the publication "Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS" which is currently available online through http://www.ai.sri.com/natural-language/projects/arpa-sls/spnl-int.html. A copy of this publication is included in an information disclosure submitted herewith, and is incorporated herein by this reference.

In an embodiment of the present invention that may be preferable for some applications, the natural language interpreter "learns" from the past usage patterns of a particular user or of groups of users. In such an embodiment, the successfully interpreted requests of users are stored, and can then be used to enhance accuracy by comparing a current request to the stored requests, thereby allowing selection of a most probable result.

### b. Constructing Navigation Queries

In step 405 request processing logic 300 identifies and selects an appropriate online data source where the desired information (in this case, current weather reports for a given city) can be found. Such selection may involve look-up in a locally stored table, or possibly dynamic searching through an online search engine, or other online search techniques. For some applications, an embodiment of the present invention may be implemented in which only access to a particular data source (such as a particular vendor's proprietary content database) is supported; in that case, step 405 may be trivial or may be eliminated entirely.

Step 406 attempts to construct a navigation query, reflecting the interpretation of step 404. This operation is preferably performed by query construction logic 330.

A "navigation query" means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information. In other words, a navigation query is constructed such that it includes whatever content and structure is required in order to access desired information electronically from a particular database or data source of interest.

For example, for many existing electronic databases, a navigation query can be embodied using a formal database query language such as Standard Query Language (SQL). For many databases, a navigation query can be constructed through a more user-friendly interactive front-end, such as a series of menus and/or interactive forms to be selected or filled in. SQL is a standard interactive and programming language for getting information from and updating a database. SQL is both an ANSI and an ISO standard. As is well known to practitioners, a Relational Database Management System (RDBMS), such as Microsoft's Access, Oracle's Oracle7, and Computer Associates' CA-OpenIngres, allow programmers to create, update, and administer a relational database. Practitioners of ordinary skill in the art will be thoroughly familiar with the notion of database navigation through structured query, and will be readily able to appreciate and utilize the existing data structures and navigational mechanisms for a given database, or to create such structures and mechanisms where desired.

In accordance with the present invention, the query constructed in step 406 must reflect the user's request as interpreted by the speech recognition engine and the NL parser in step 404. In embodiments of the present invention wherein data source 110 (or 210 in the corresponding embodiment of Figure 2) is a structured relational database or the like, step 406 of the present invention may entail constructing an

15

appropriate Structured Query Language (SQL) query or the like, or automatically filling out a front-end query form, series of menus or the like, as described above.

In many existing Internet (and Intranet) applications, an online electronic data source is accessible to users only through the medium of interaction with a so-called Common Gateway Interface (CGI) script. Typically the user who visits a web site of this nature must fill in the fields of an online interactive form. The online form is in turn linked to a CGI script, which transparently handles actual navigation of the associated data source and produces output for viewing by the user's web browser. In other words, direct user access to the data source is not supported, only mediated access through the form and CGI script is offered.

For applications of this nature, an advantageous embodiment of the present invention "scrapes" the scripted online site where information desired by a user may be found in order to facilitate construction of an effective navigation query. For example, suppose that a user's spoken natural language request is: "What's the weather in Miami?" After this request is received at step 402 and interpreted at step 404, assume that step 405 determines that the desired weather information is available online through the medium of a CGI-scripted interactive form. Step 406 is then preferably carried out using the expanded process diagrammed in Figure 5. In particular, at sub-step 520, query construction logic 330 electronically "scrapes" the online interactive form, meaning that query construction logic 330 automatically extracts the format and structure of input fields accepted by the online form. At sub-step 522, a navigation query is then constructed by instantiating (filling in) the extracted input format -- essentially an electronic template -- in a manner reflecting the user's request for information as interpreted in step 404. The flow of control then returns to step 407 of Figure 4. Ultimately, when the query thus constructed by scraping is used to navigate the online data source in step 408, the query effectively initiates the same scripted response as if a human user had visited the online site and had typed appropriate entries into the input fields of the online form.

In the embodiment just described, scraping step 520 is preferably carried out with the assistance of an online extraction utility such as WebL. WebL is a scripting language for automating tasks on the World Wide Web. It is an imperative,

16

interpreted language that has built-in support for common web protocols like HTTP and FTP, and popular data types like HTML and XML. WebL's implementation language is Java, and the complete source code is available from Compaq. In addition, step 520 is preferably performed dynamically when necessary -- in other words, on-the-fly in response to a particular user query -- but in some applications it may be possible to scrape relatively stable (unchanging) web sites of likely interest in advance and to cache the resulting template information.

It will be apparent, in light of the above teachings, that preferred embodiments of the present invention can provide a spoken natural language interface atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the linear browsing architecture or other artifacts of an existing menu/text/click navigation system. For example, users of an appropriate embodiment of the present invention for a video-on-demand application can directly speak the natural request: "Show me the movie 'Unforgiven'" -- instead of walking step-by-step through a typically linear sequence of genre/title/actor/director menus, scrolling and selecting from potentially long lists on each menu, or instead of being forced to use an alphanumeric keyboard that cannot be as comfortable to hold or use as a lightweight remote control. Similarly, users of an appropriate embodiment of the present invention for a web-surfing application in accordance with the process shown in Figure 5 can directly speak the natural request: "Show me a one-month price chart for Microsoft stock" -- instead of potentially having to navigate to an appropriate web site, search for the right ticker symbol, enter/select the symbol, and specify display of the desired one-month price chart, each of those steps potentially involving manual navigation and data entry to one or more different interaction screens. (Note that these examples are offered to illustrate some of the potential benefits offered by appropriate embodiments of the present invention, and not to limit the scope of the invention in any respect.)

c. Error Correction

Several problems can arise when attempting to perform searches based on spoken natural language input. As indicated at decision step 407 in the process of Figure 4, certain deficiencies may be identified during the process of query construction, before search of the data source is even attempted. For example, the user's request may fail to specify enough information in order to construct a navigation query that is specific enough to obtain a satisfactory search result. For example, a user might orally request "what's the weather?" whereas the national online data source identified in step 405 and scraped in step 520 might require specifying a particular city.

Additionally, certain deficiencies and problems may arise following the navigational search of the data source at step 408, as indicated at decision step 409 in Figure 4. For example, with reference to a video-on-demand application, a user may wish to see the movie "Unforgiven", but perhaps the user can't recall name of the film, but knows it was directed by and starred actor Clint Eastwood. A typical video-on-demand database might indeed be expected to allow queries specifying the name of a leading actor and/or director, but in the case of this query -- as in many cases -- that will not be enough to narrow the search to a single film, and additional user input in some form is required.

In the event that one or more deficiencies in the user's spoken request, as processed, result in the problems described, either at step 407 or 409, some form of error handling is in order. A straightforward, crude technique might be for the system to respond simply *"input not understood / insufficient; please try again."* However, that approach will likely result in frustrated users, and is not optimal or even acceptable for most applications. Instead, a preferred technique in accordance with the present invention handles such errors and deficiencies in user input at step 412, whether detected at step 407 or step 409, by soliciting additional input from the user in a manner taking advantage of the partial construction already performed and via user interface modalities in addition to spoken natural language ("multi-modality"). This supplemental interaction is preferably conducted through client display device 112 (202, in the embodiment of Figure 2), and may include textual, graphical, audio and/or video media. Further details and examples are provided below. Query refinement logic 340 preferably carries out step 412. The additional input received from the user is fed into and augments interpreting step 404, and query construction step 406 is likewise repeated with the benefit of the augmented interpretation. These operations, and subsequent navigation step 408, are preferably repeated until no

18

remaining problems or deficiencies are identified at decision points 407 or 409. Further details and examples for this query refinement process are provided immediately below.

Consider again the example in which the user of a video-on-demand application wishes to see "Unforgiven" but can only recall that it was directed by and starred Clint Eastwood. First, it bears noting that using a prior art navigational interface, such as a conventional menu interface, will likely be relatively tedious in this case. The user can proceed through a sequence of menus, such as Genre (select "western"), Title (skip), Actor ("Clint Eastwood"), and Director ("Clint Eastwood"). In each case --especially for the last two items -- the user would typically scroll and select from fairly long lists in order to enter his or her desired name, or perhaps use a relatively couch-unfriendly keypad to manually type the actor's name twice.

Using a preferred embodiment of the present invention, the user instead speaks aloud, holding remote control microphone 102, "I want to see that movie starring and directed by Clint Eastwood. Can't remember the title." At step 402 the voice data is received. At step 404 the voice data is interpreted. At step 405 an appropriate online data source is selected (or perhaps the system is directly connected to a proprietary video-on-demand provider). At step 406 a query is automatically constructed by the query construction logic 330 specifying "Clint Eastwood" in both the actor and director fields. Step 407 detects no obvious problems, and so the query is electronically submitted and the data source is navigated at step 408, yielding a list of several records satisfying the query (e.g., "Unforgiven", "True Crime", "Absolute Power", etc.). Step 409 detects that additional user input is needed to further refine the query in order to select a particular film for viewing.

At that point, in step 412 query refinement logic 340 might preferably generate a display for client display device 112 showing the (relatively short) list of film titles that satisfy the user's stated constraints. The user can then preferably use a relatively convenient input modality, such as buttons on the remote control, to select the desired title from the menu. In a further preferred embodiment, the first title on the list is highlighted by default, so that the user can simply press an "OK" button to choose that selection. In a further preferred feature, the user can mix input modalities by speaking a response like "I want number one on the list." Alternatively, the user can preferably say, "Let's see Unforgiven," having now been reminded of the title by the menu display.

Utilizing the user's supplemental input, request processing logic 300 iterates again through steps 404 and 406, this time constructing a fully-specified query that specifically requests the Eastwood film "Unforgiven." Step 408 navigates the data source using that query and retrieves the desired film, which is then electronically transmitted in step 410 from network server 108 to client display device 112 via communications network 106.

Now consider again the example in which the user of a web surfing application wants to know his or her local weather, and simply asks, "what's the weather?" At step 402 the voice data is received. At step 404 the voice data is interpreted. At step 405 an online web site providing current weather information for major cities around the world is selected. At step 406 and sub-step 520, the online site is scraped using a WebL-style tool to extract an input template for interacting with the site. At sub-step 522, query construction logic 330 attempts to construct a navigation query by instantiating the input template, but determines (quite rightly) that a required field -- name of city -- cannot be determined from the user's spoken request as interpreted in step 404. Step 407 detects this deficiency, and in step 412 query refinement logic 340 preferably generates output for client display device 112 soliciting the necessary supplemental input. In a preferred embodiment, the output might display the name of the city where the user is located highlighted by default. The user can then simply press an "OK" button -- or perhaps mix modalities by saying "yes, exactly" -- to choose that selection. A preferred embodiment would further display an alphabetical scrollable menu listing other major cities, and/or invite the user to speak or select the name of the desired city.

Here again, utilizing the user's supplemental input, request processing logic 300 iterates through steps 404 and 406. This time, in performing sub-step 520, a cached version of the input template already scraped in the previous iteration might preferably be retrieved. In sub-step 522, query construction logic 330 succeeds this time in instantiating the input template and constructing an effective query, since the desired city has now been clarified. Step 408 navigates the data source using that query and retrieves the desired weather information, which is then electronically transmitted in step 410 from network server 108 to client display device 112 via communications network 106.

It is worth noting that in some instances, there may be details that are not explicitly provided by the user, but that query construction logic 330 or query refinement logic 340 may preferably deduce on their own through reasonable assumptions, rather than requiring the use to provide explicit clarification. For example, in the example previously described regarding a request for a weather report, in some applications it might be preferable for the system to simply assume that the user means a weather report for his or her home area and to retrieve that information, if the cost of doing so is not significantly greater than the cost of asking the user to clarify the query. Making such an assumption might be even more strongly justified in a preferred embodiment, as described earlier, where user histories are tracked, and where such history indicates that a particular user or group of users typically expect local information when asking for a weather forecast. At any rate, in the event such an assumption is made, if the user actually intended to request the weather for a different city, the user would then need to ask his or her question again. It will be apparent to practitioners, in light of the above teachings, that the choice of whether to program query construction logic 330 and query refinement logic 340 to make make particular assumptions will typically involve trade-offs involving user conveience that can be assessed in the context of specific applications.

### 3. Open Agent Architecture (OAA®)

Open Agent Architecture™ (OAA®) is a software platform, developed by the assignee of the present invention, that enables effective, dynamic collaboration among communities of distributed electronic agents. OAA is described in greater detail in co-pending U.S. Patent Application No. 09/225,198, which has been incorporated herein by reference. Very briefly, the functionality of each client agent is made available to the agent community through registration of the client agent's capabilities with a facilitator. A software "wrapper" essentially surrounds the underlying application program performing the services offered by each client. The common infrastructure for constructing agents is preferably supplied by an *agent library*. The agent library is preferably accessible in the runtime environment of several different programming languages. The agent library preferably minimizes the effort required to construct a new system and maximizes the ease with which legacy systems can be "wrapped" and made compatible with the agent-based architecture of the present invention. When invoked, a client agent makes a connection to a facilitator, which is known as its *parent facilitator*. Upon connection, an agent registers with its parent facilitator a specification of the capabilities and services it can provide, using a high-level, declarative Interagent Communication Language (*"ICL"*) to express those capabilities. Tasks are presented to the facilitator in the form of ICL goal expressions. When a facilitator determines that the registered capabilities of one of its client agents will help satisfy a current goal or sub-goal thereof, the facilitator delegates that sub-goal to the client agent in the form of an ICL request. The client agent processes the request and returns answers or information to the facilitator. In processing a request, the client agent can use *ICL* to request services of other agents, or utilize other infrastructure services for collaborative work. The facilitator coordinates and integrates the results received from different client agents on various sub-goals, in order to satisfy the overall goal.

OAA provides a useful software platform for building systems that integrate spoken natural language as well as other user input modalities. For example, see the above-referenced co-pending patent application, especially Figure 13 and the corresponding discussion of a "multi-modal maps" application, and Figure 12 and the

22

corresponding discussion of a "unified messaging" application. Another example is the InfoWiz interactive information kiosk developed by the assignee and described in the document entitled "InfoWiz: An Animated Voice Interactive Information System" available online at http://www.ai.sri.com/~oaa/applications.html. A copy of the InfoWhiz document is provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference. A further example is the "CommandTalk" application developed by the assignee for the U.S. military, as described online at http://www.ai.sri.com/~lesaf/commandtalk.html and in the following publications, copies of which are provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference:

- "CommandTalk: A Spoken-Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, DC, pp. 1-7, Association for Computational Linguistics
- "The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty-Seventh Annual Meeting of the ACL", pp. 183-190, University of Maryland, College Park, MD, Association for Computational Linguistics
- "Interpreting Language in Context in CommandTalk", 1999, by John Dowding and Elizabeth Owen Bratt and Sharon Goldwater, in "Communicative Agents: The Use of Natural Language in Embodied Systems", pp. 63-67, Association for Computing Machinery (ACM) Special Interest Group on Artificial Intelligence (SIGART), Seattle, WA

For some applications and systems, OAA can provide an advantageous platform for constructing embodiments of the present invention. For example, a representative application is now briefly presented, with reference to Figure 6. If the statement "show me movies starring John Wayne" is spoken into the voice input device, the voice data for this request will be sent by UI agent 650 to facilitator 600, which in turn will ask natural language (NL) agent 620 and speech recognition agent 610 to interpret the query and return the interpretation in *ICL* format. The resulting *ICL* goal expression is then routed by the facilitator to appropriate agents -- in this case, video-on-demand database agent 640 -- to execute the request. Video database agent 640 preferably includes or is coupled to an appropriate embodiment of query construction logic 330 and query refinement logic 340, and may also issue ICL requests to facilitator 600 for additional assistance -- e.g., display of menus and capture of additional user input in the event that query refinement is needed -- and facilitator 600 will delegate such requests to appropriate client agents in the community. When the desired video content is ultimately retrieved by video database agent 640, UI agent 650 is invoked by facilitator 600 to display the movie.

Other spoken user requests, such as a request for the current weather in New York City or for a stock quote, would eventually lead facilitator to invoke web database agent 630 to access the desired information from an appropriate Internet site. Here again, web database agent 630 preferably includes or is coupled to an appropriate embodiment of query construction logic 330 and query refinement logic 340, including a scraping utility such as WebL. Other spoken requests, such as a request to view recent emails or access voice mail, would lead the facilitator to invoke the appropriate email agent 660 and/or telephone agent 680. A request to record a televised program of interest might lead facilitator 600 to invoke web database agent 630 to return televised program schedule information, and then invoke VCR controller agent 680 to program the associated VCR unit to record the desired television program at the scheduled time.

Control and connectivity embracing additional electronic home appliances (e.g., microwave oven, home surveillance system, etc.) can be integrated in comparable fashion. Indeed, an advantage of OAA-based embodiments of the present invention, that will be apparent to practitioners in light of the above teachings and in light of the teachings disclosed in the cited co-pending patent applications, is the relative ease and flexibility with which additional service agents can be plugged into the existing platform, immediately enabling the facilitator to respond dynamically to spoken natural language requests for the corresponding services.

24

## 4. Further Embodiments and Equivalents

While the present invention has been described in terms of several preferred embodiments, there are many alterations, permutations, and equivalents that may fall within the scope of this invention. It should also be noted that there are many alternative ways of implementing the methods and apparatuses of the present invention. It is therefore intended that the following appended claims be interpreted as including all such alterations, permutations, and equivalents as fall within the true spirit and scope of the present invention.

## CLAIMS

***What is claimed is:***

1. A method for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken natural language ("NL") request for desired information from the user;

(b) rendering an interpretation of the spoken natural language request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a modality different than the original request;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

2. The method of claim 1, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and an NL parser.

3. The method of claim 1, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

4. The method of claim 3, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

5. The method of claim 1, wherein the navigation query is constructed in the format of a database query language.

6. The method of claim 1, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

7. The method of claim 1, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

8. The method of claim 1, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

9. The method of claim 8, wherein the deficiencies include unresolved words of the spoken NL request.

10. The method of claim 8, wherein the deficiencies include one or more required elements of the navigational query ~~not~~ determinable from the interpretation of the spoken NL request.

11. The method of claim 1, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

12. The method of claim 11, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

13. The method of claim 11, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

14. The method of claim 1, wherein the input modality of step (d) includes selecting from a displayed option menu.

15. The method of claim 14, wherein the act of selecting from the displayed option menu is performed by speaking.

16. The method of claim 1, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

17. The method of claim 1, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

18. The method of claim 1, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

19. A system for utilizing spoken natural language to navigate an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken natural language ("NL") request for desired information from the user;

(b) spoken language processing logic, operable to render an interpretation of the spoken natural language request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken natural language request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a modality different than the original request;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

20. The system of claim 19, wherein the spoken language processing logic includes speech recognition logic and an NL parsing logic for deriving linguistic information.

21. The system of claim 19, wherein the spoken language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

22. The system of claim 21, wherein the spoken language processing logic dynamically scrapes the online scripted interface.

23. The system of claim 19, wherein the query construction logic constructs the query in the format of a database query language.

24. The system of claim 19, wherein at least a portion of the spoken language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

25. The system of claim 19, wherein at least a portion of the spoken language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

26. The system of claim 19, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

27. The system of claim 26, wherein the deficiencies include unresolved words of the spoken NL request.

28. The system of claim 26, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

29. The system of claim 19, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

30. The system of claim 29, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

31. The system of claim 29, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

32. The system of claim 19, wherein the user interaction logic displays an option menu.

33. The system of claim 32, wherein the act of selecting from the displayed option menu is performed by speaking.

34. The system of claim 19, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

35. The system of claim 19, wherein the electronic data source stores multimedia content including at least ~~one of~~ video content and audio content.

36. The system of claim 19, wherein the display device receives data from the electronicdata source on the network servers via a communications box.

37. The system of claim 19, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

38. An computer program embodied on a computer readable medium for utilizing spoken natural language for navigating an electronic data source, the

electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken natural language ("NL") request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken natural language request;

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;

(e) a code segment that refines the navigation query, based upon the additional input;

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

39. The computer program of claim 38, further comprising a code segment that derives linguistic information by using a speech recognition engine and an NL parser.

40. The computer program of claim 38, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

41. The computer program of claim 40, further comprising a code segment that dynamically scrapes the online scripted interface.

42. The computer program of claim 38, wherein the navigation query is constructed in the format of a database query language.

43. The computer program of claim 38, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

44. The computer program of claim 38, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.

45. The computer program of claim 38, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.

46. The computer program of claim 45, wherein the deficiencies include unresolved words of the spoken NL request.

47. The computer program of claim 45, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

48. The computer program of claim 38, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.

49. The computer program of claim 48, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

50. The computer program of claim 48, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

51. The computer program of claim 38, wherein code segment that solicits additional input displays an option menu.

52. The computer program of claim 51, wherein the act of selecting from the displayed option menu is performed by speaking.

~~53. The computer program of claim 38, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.~~

~~54. The computer program of claim 38, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.~~

~~55. The computer program of claim 38, wherein the electronic data source stores multimedia content including at least one of video content and audio content.~~

add B4

# NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

## ABSTRACT OF THE INVENTION

Sub B1

A system, method, and article of manufacture are provided for navigating an electronic data source by means of spoken natural language. When a spoken natural language input request is received from a user, it is interpreted. Additional input is solicited from the user in a modality different than the original request and used to refine the navigation query. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources.

1

|  |  |  |
|---|---|---|
| Post-it® Fax Note 7671 | Date 6/29/00 | # of pages ▶ 7 |
| To Domenic Rotab | From K. Elbisni |  |
| Co./Dept. | Co. SRI Intl. |  |
| Phone # 408-971-4660 | Phone # 650-859-6631 |  |
| Fax # 408-971-4660 | Fax # 650.859-6420 |  |

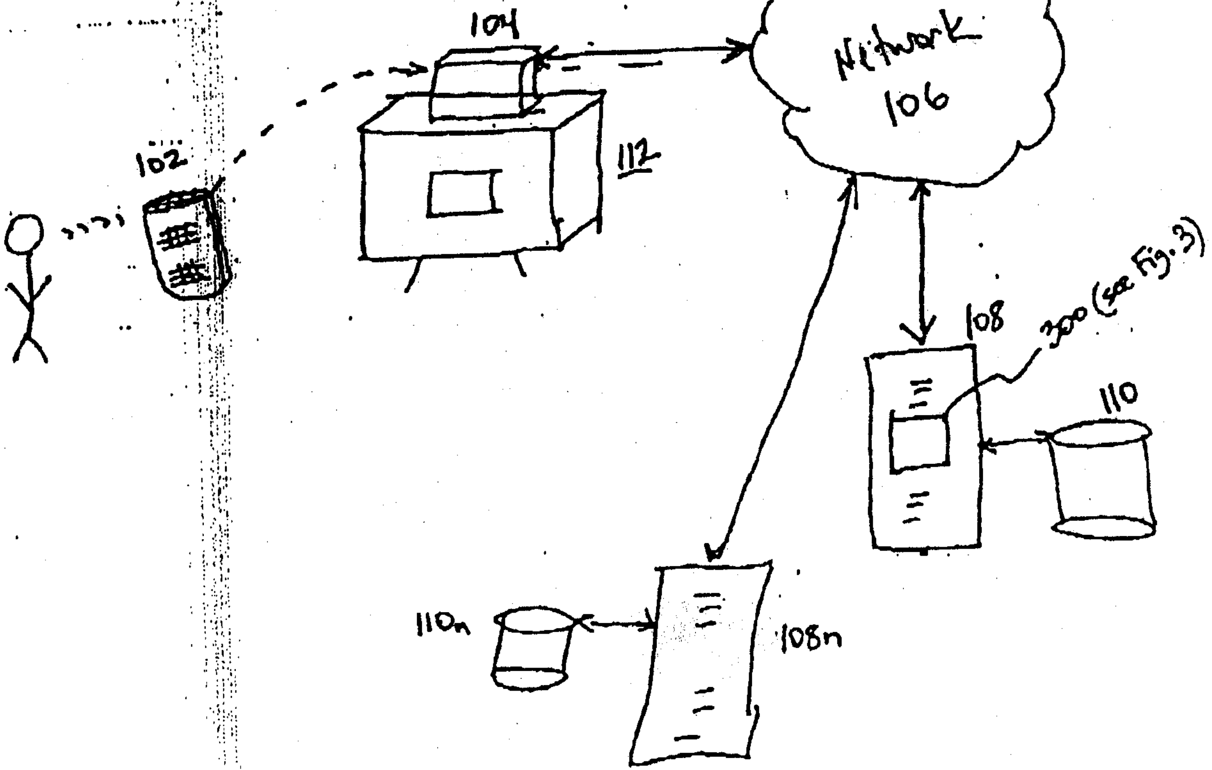


Fig. 1a

09608237 .062900

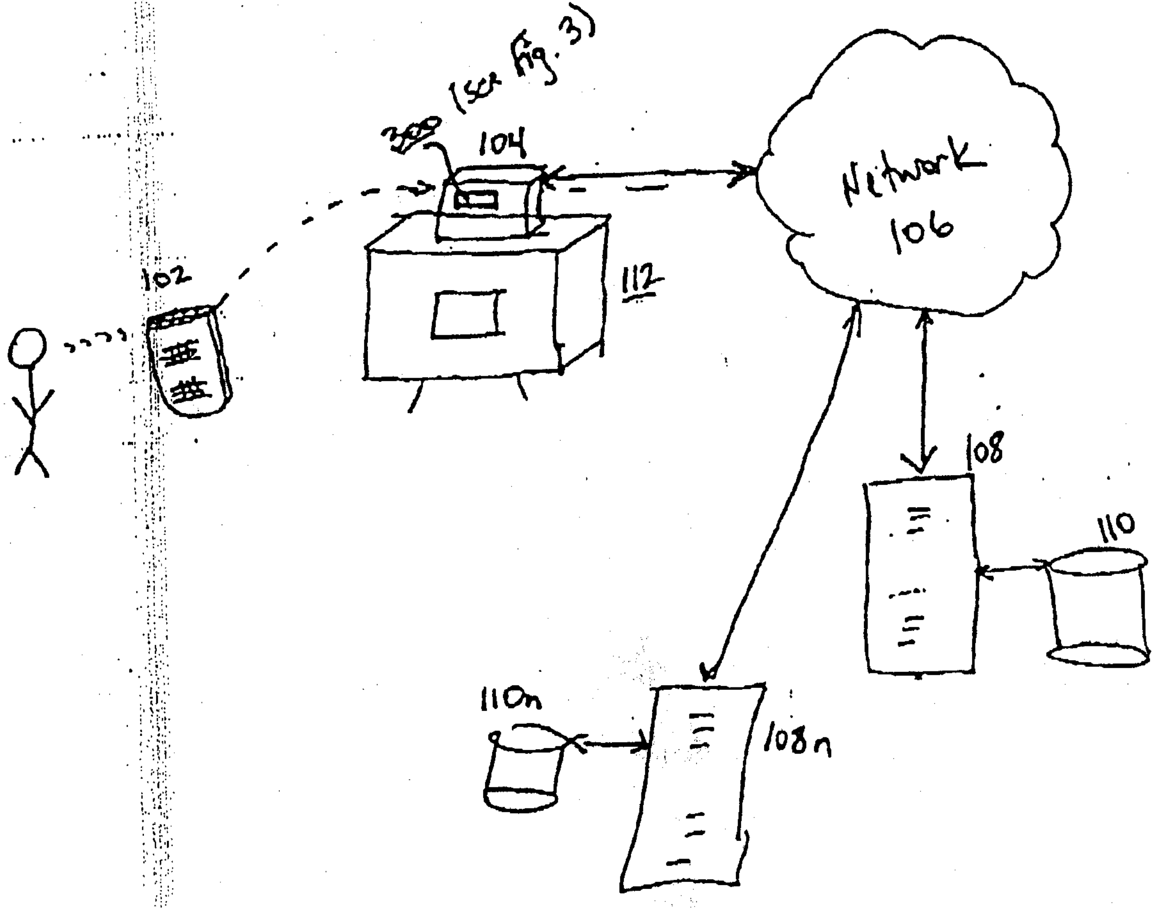


Fig. 1b

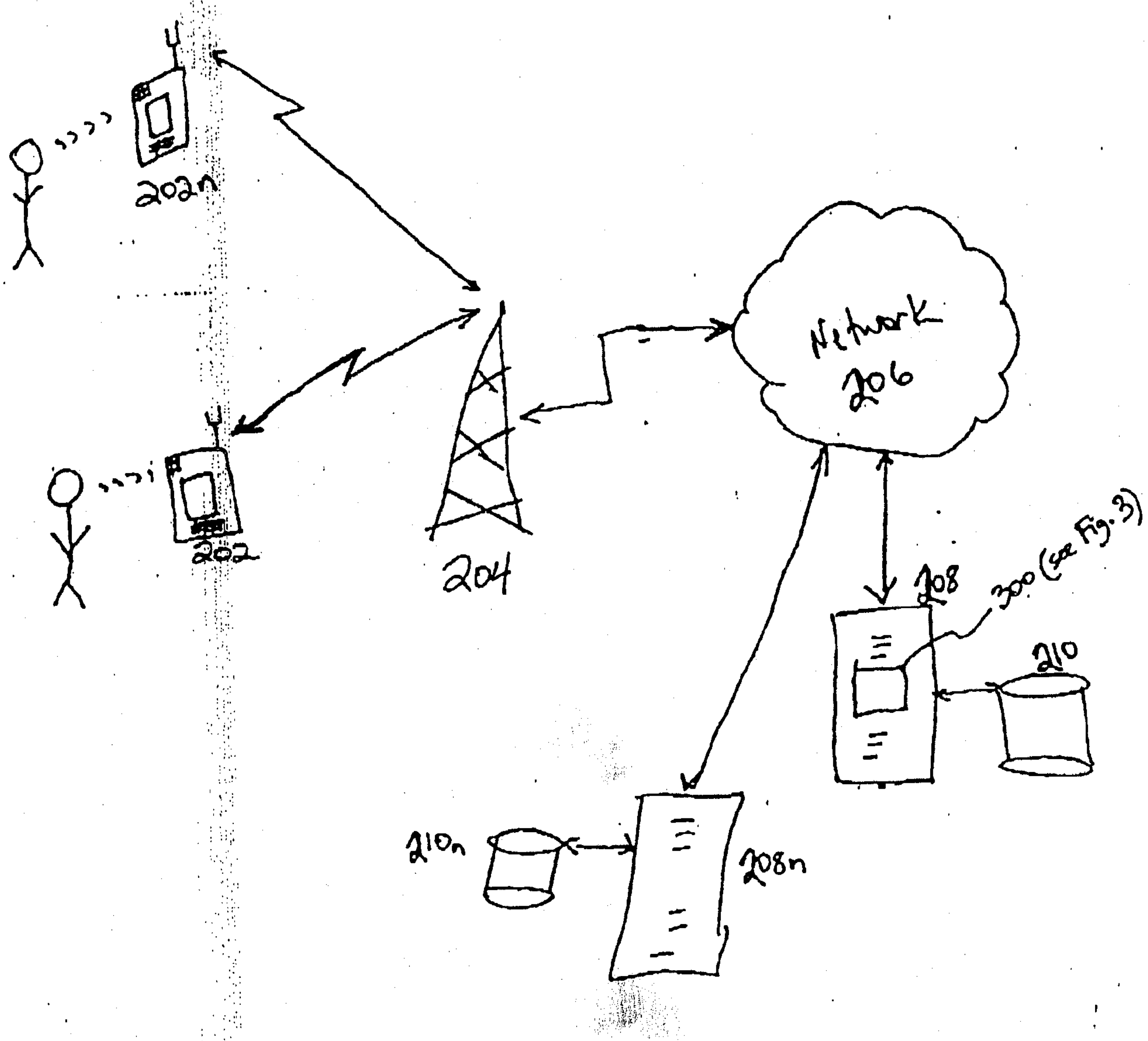


Fig. 2

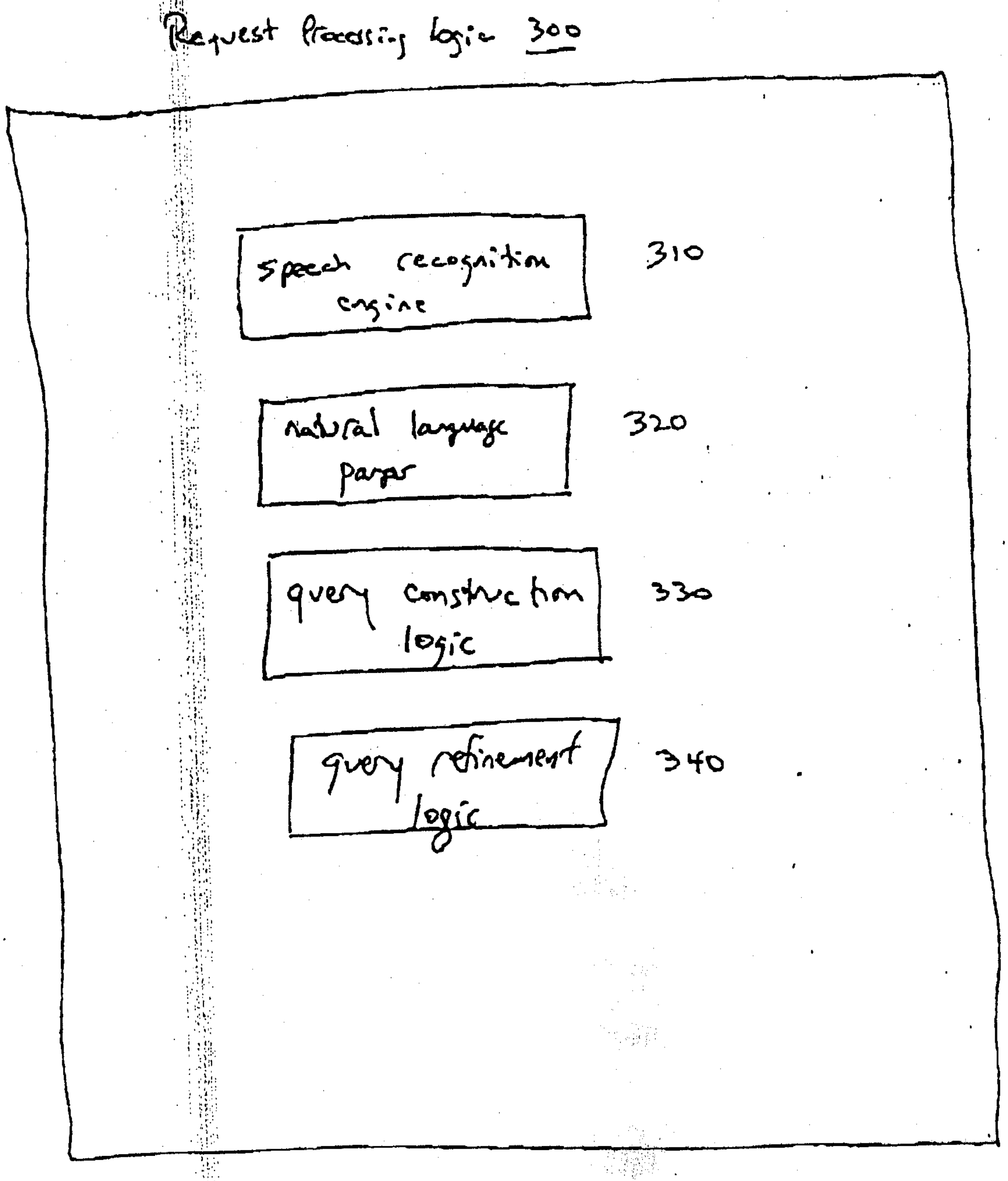

Request Processing logic 300
speech recognition engine
310
natural language parser
320
query construction logic
330
query refinement logic
340

Fig. 3

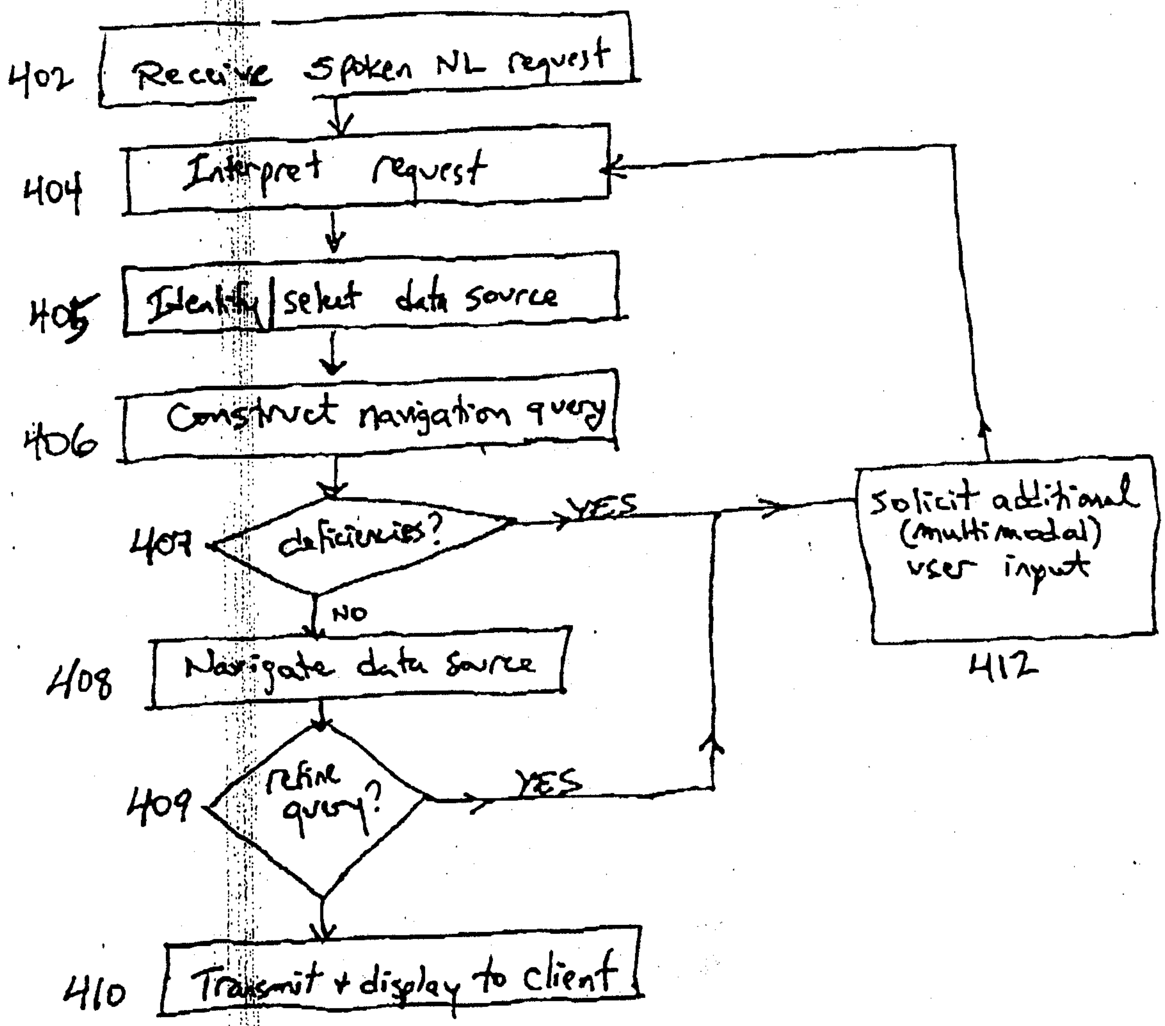

402
Receive spoken NL request
404
Interpret request
405
Identify/select data source
406
Construct navigation query
407
deficiencies?
YES
NO
408
Navigate data source
409
refine query?
YES
410
Transmit + display to client
Solicit additional (multimodal) user input
412

Fig. 4

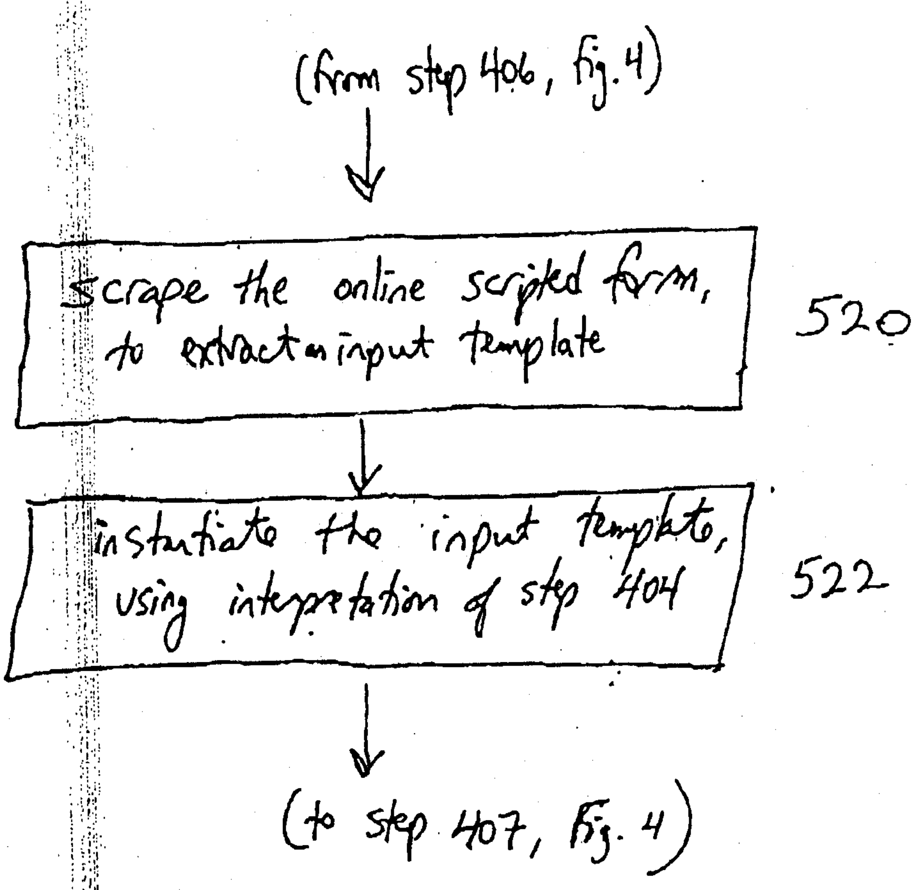

(from step 406, fig. 4)
scrape the online scripted form, to extract an input template
520
instantiate the input template, using interpretation of step 404
522
(to step 407, Fig. 4)

Fig. 5

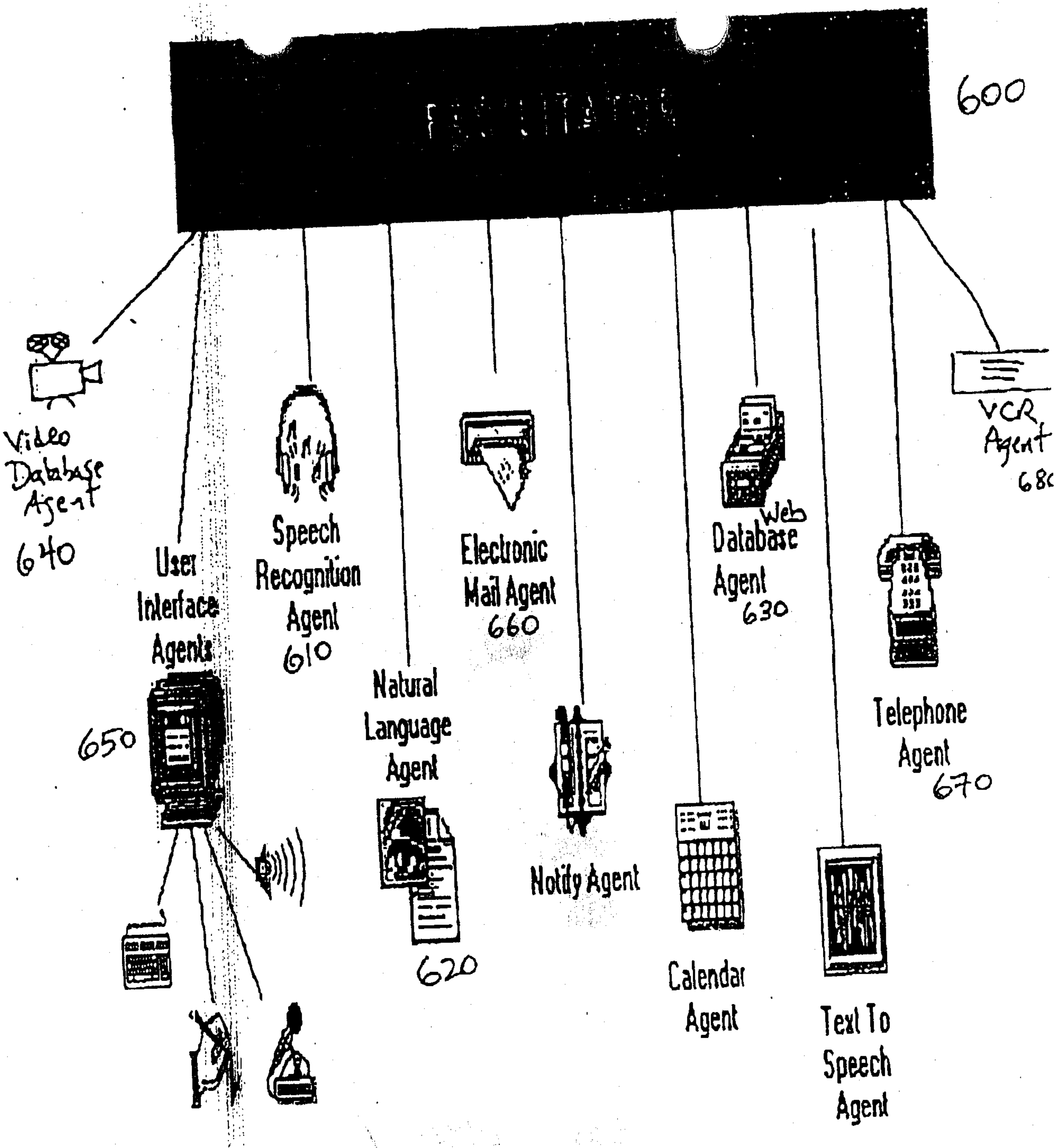

FACILITATOR
600
Video Database Agent
640
User Interface Agents
650
Speech Recognition Agent
610
Natural Language Agent
620
Electronic Mail Agent
660
Notify Agent
Calendar Agent
Web Database Agent
630
Text To Speech Agent
Telephone Agent
670
VCR Agent
68

Fig. 6

**PRINT OF DRAWINGS**
**AS ORIGINALLY FILEI**

⑦

| | | | |
|---|---|---|---|
| Post-it® Fax Note 7671 | | Date 6/29/00 | # of pages ▶ 7 |
| To Domenic Kotab | | From K. Eloisni | |
| Co./Dept. | | Co. SRI Intl. | |
| Phone # 408-971-4660 | | Phone # 650-859-6631 | |
| Fax # 408-971-4660 | | Fax # 650.859-6420 | |

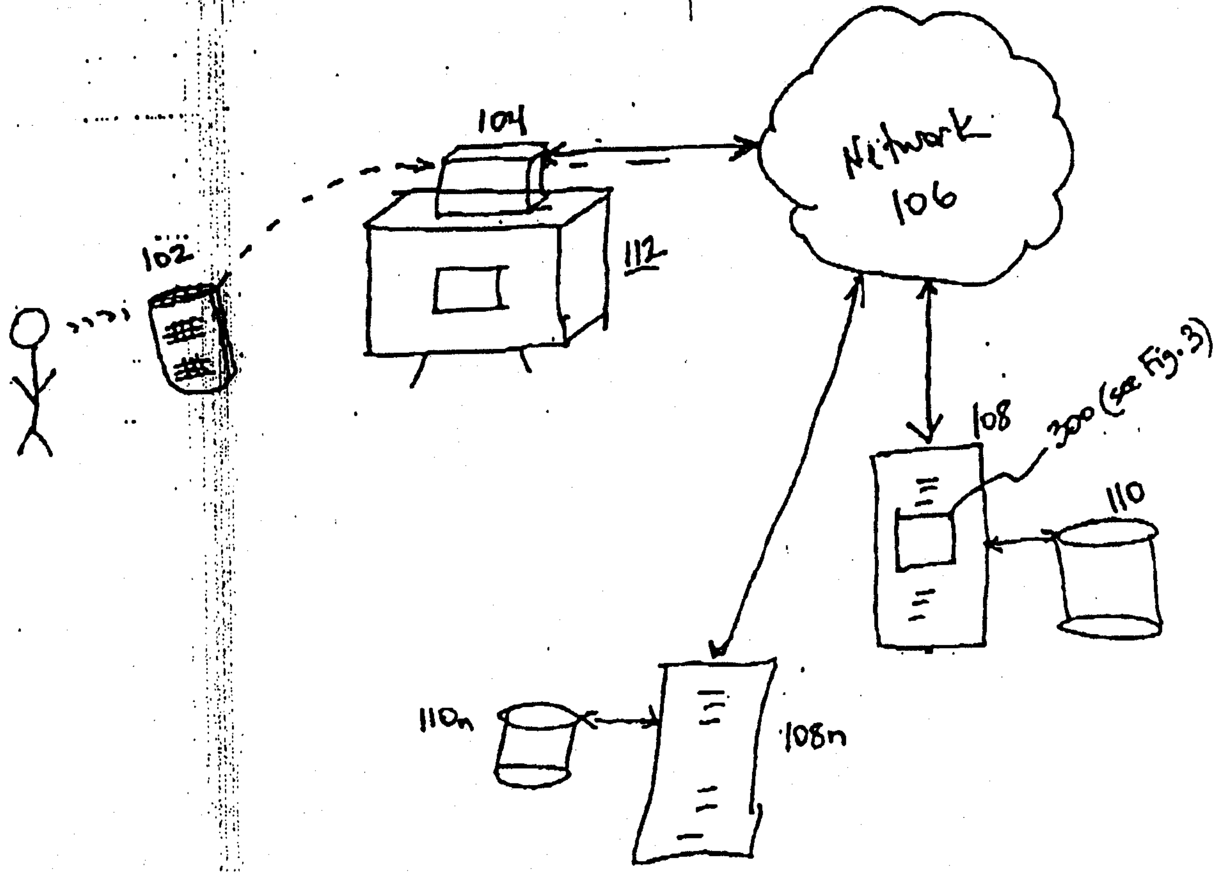


Fig. 1a

PRINT OF DRAWINGS
AS ORIGINALLY FILED

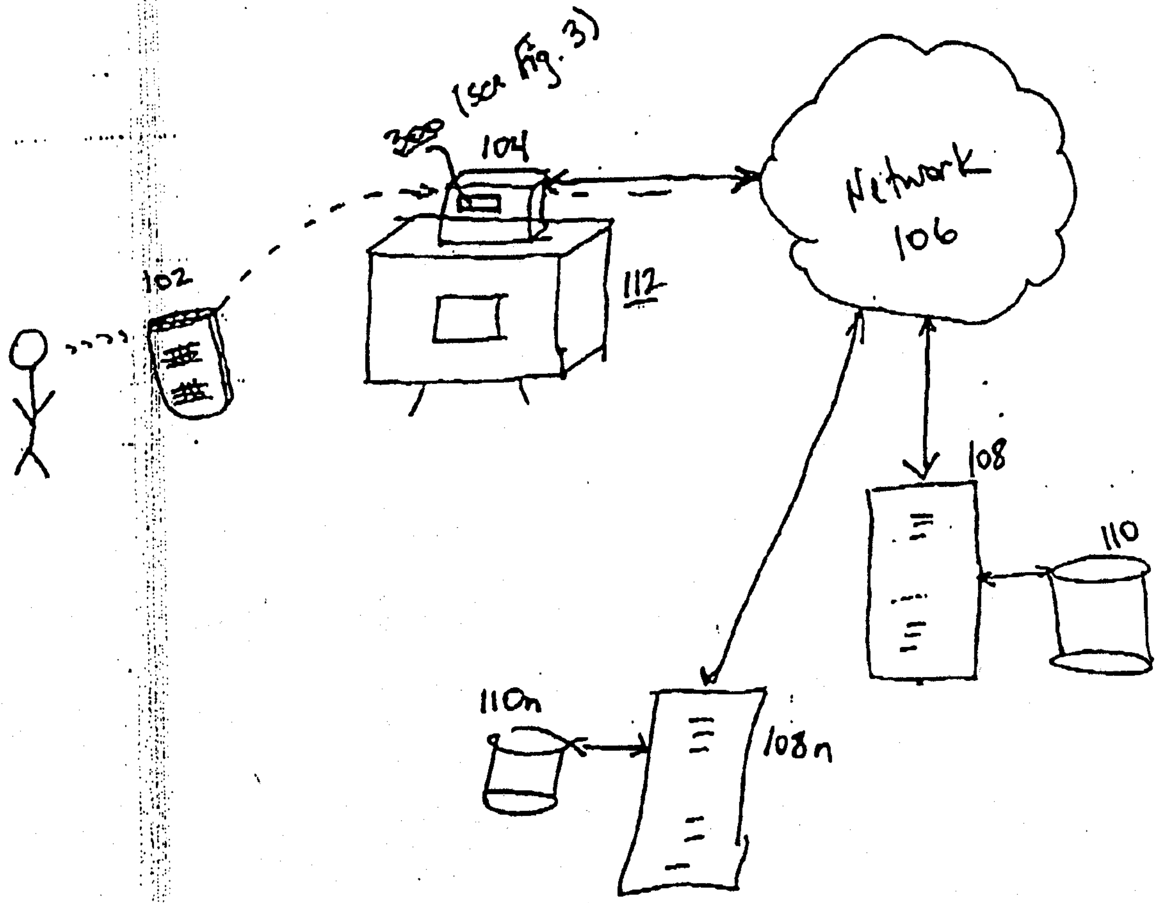


Fig. 1b

PRINT OF DRAWINGS
AS ORIGINALLY FILEI

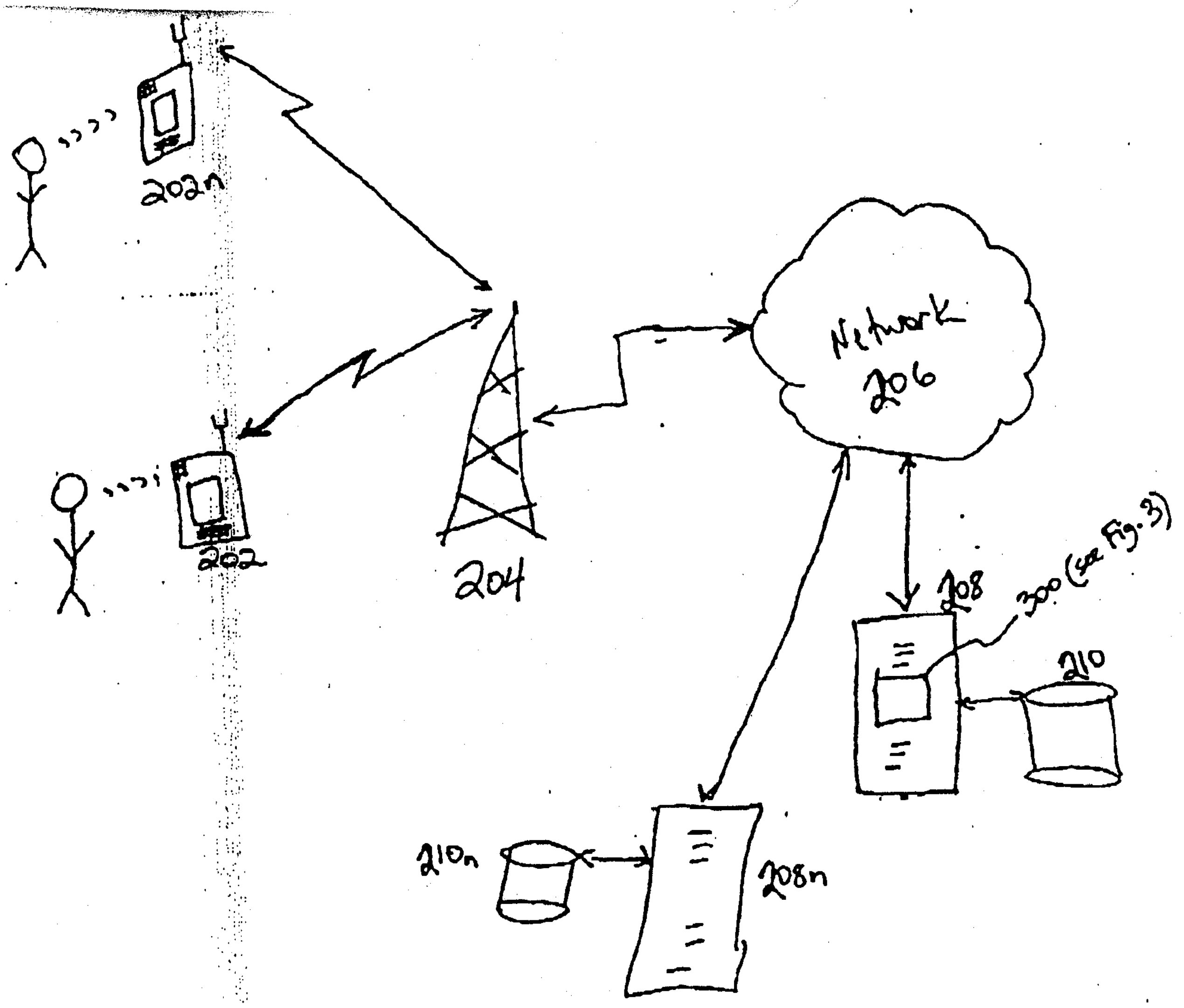


Fig. 2

PRINT OF DRAWINGS
AS ORIGINALLY FILED

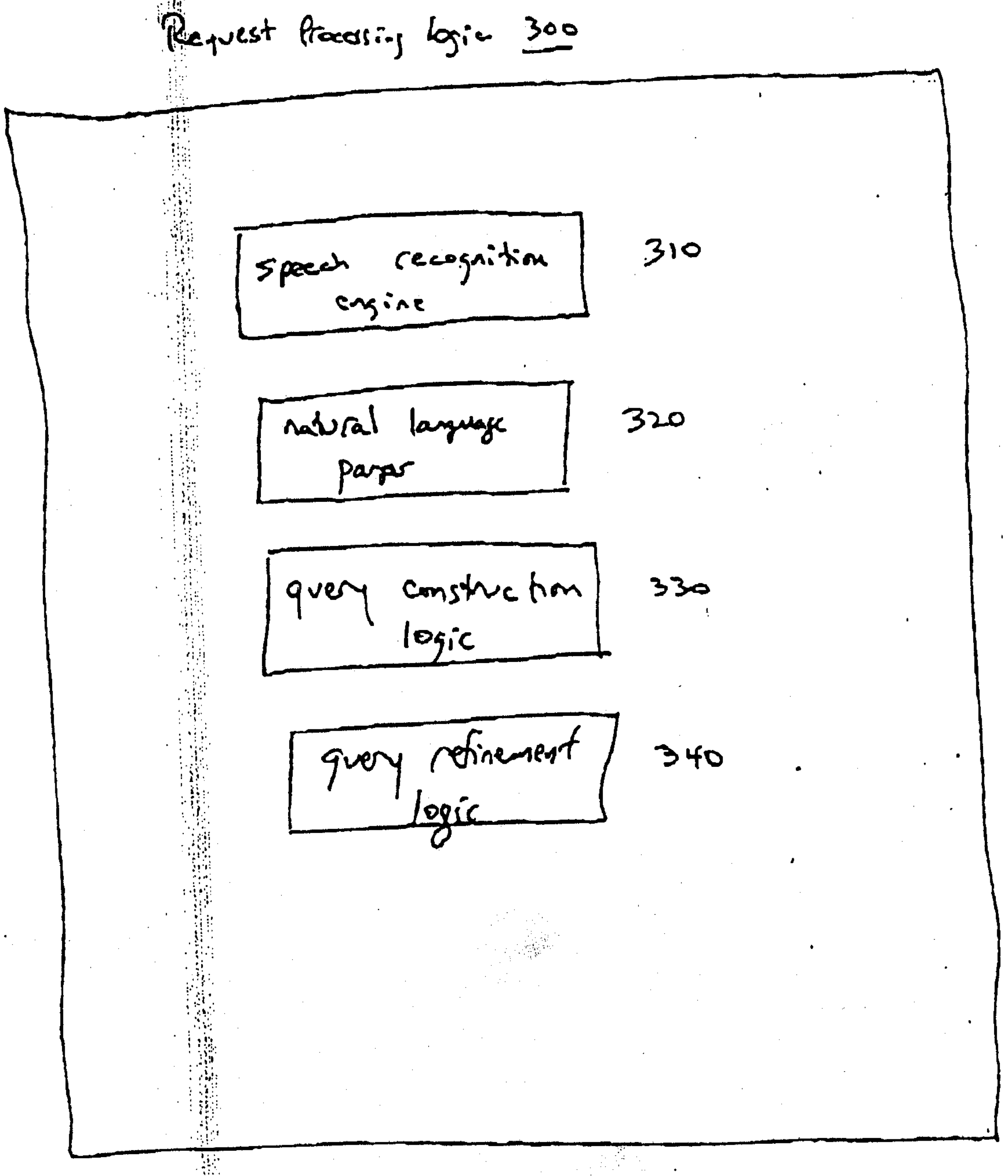


Fig. 3

PRINT OF DRAWINGS
AS ORIGINALLY FILED

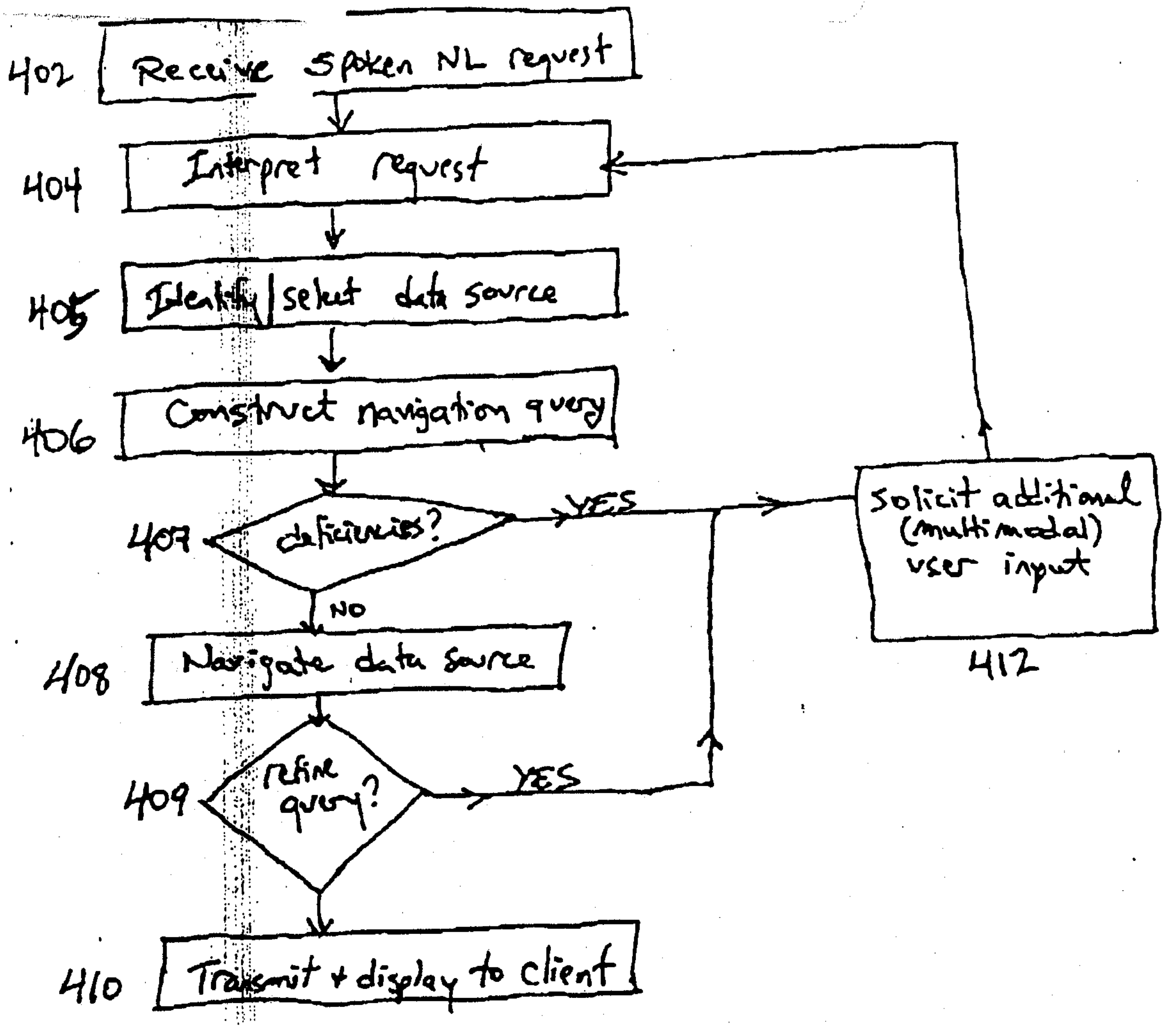


Fig. 4

PRINT OF DRAWINGS
AS ORIGINALLY FILED

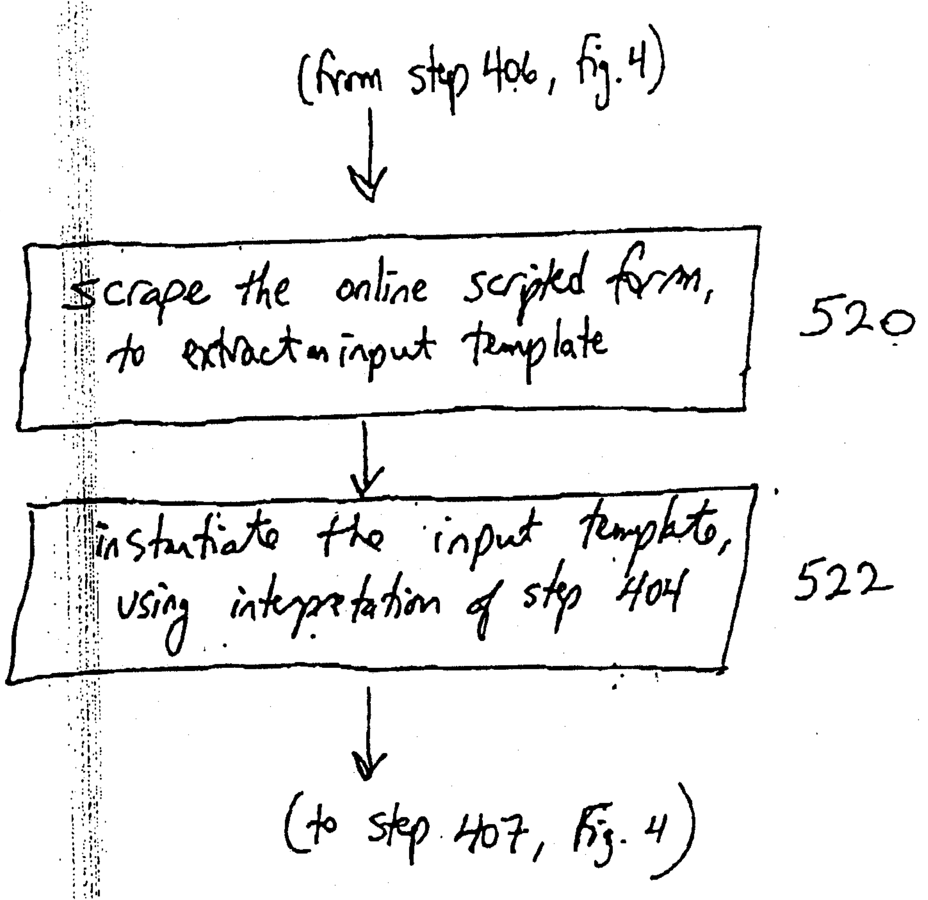


Fig. 5

PRINT OF DRAWINGS
AS ORIGINALLY FILED

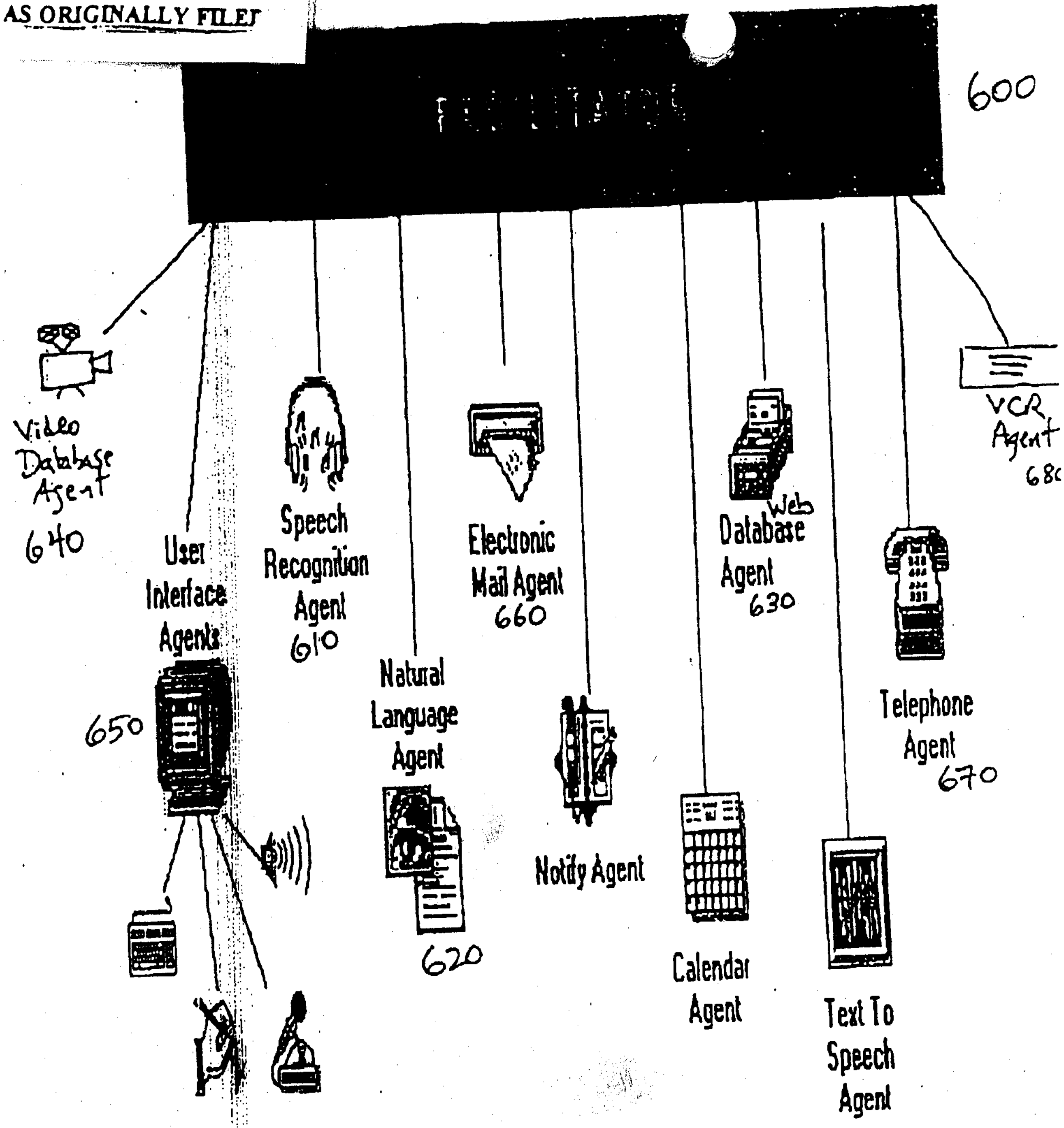


Fig. 6

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

| APPLICATION NUMBER | FILING/RECEIPT DATE | FIRST NAMED APPLICANT | ATTORNEY DOCKET NUMBER |
|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRI1p037B |

**FORMALITIES LETTER**

*OC000000005370740*

Kevin J Zilka
P O Box 721030
San Jose, CA 95172-1030

Date Mailed: 09/01/2000

## NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

### FILED UNDER 37 CFR 1.53(b)

***Filing Date Granted***

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given TWO MONTHS from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is missing.
  *A properly signed oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date, is required.*
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of $65 for a small entity in compliance with 37 CFR 1.27, must be submitted with the missing items identified in this letter.

- **The balance due by applicant is $ 65.**

---

***A copy of this notice MUST be returned with the reply.***

M. Middleton
Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 3 - OFFICE COPY

OIPE JC131
NOV 0 2 2000
PATENT & TRADEMARK OFFICE

**PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the application of )
) Examiner: Not Assigned
Halverson et al. )
) Art Unit: 2741
Application No. 09/608,872 )
) Atty. Docket No. SRI1P037B
Filed: June 30, 2000 )
) Date: October 30, 2000
For: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT )

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231 on October 30, 2000.

Signed: Julie A. Curts
Julie A. Curts

RESPONSE TO NOTICE TO FILE MISSING PARTS

Assistant Commissioner for Patents
**Box: Missing Parts**
Washington, D.C. 20231

Sir:

In response to the Notice to File Missing Parts of Application--Filing Date Granted dated September 1, 2000, Applicants hereby attach an original executed Declaration and Power of Attorney, and the copy of the Notice to be returned with this response.

Applicants are also attaching Check No. 238 for $65.00 in payment of the surcharge fee. The Commissioner is authorized to charge any other fees that may be due to our Deposit Account No. 50-1351 (Order No. SRI1P037B). A copy of this sheet is enclosed for this purpose.

Respectfully submitted,
SILICON VALLEY IP LAW GROUP

Kevin J. Zilka
Reg. No 41,429

P.O. Box 721030
San Jose, CA 95172-1030
(408) 505-5100

Attorney Docket No. SRI1P037B

NOV 0 2 2000
PATENT & TRADEMARK OFFICE

# DECLARATION AND POWER OF ATTORNEY FOR ORIGINAL U.S. PATENT APPLICATION

Attorney's Docket No. SRI1P037

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: NAVIGATING NETWORK-BASED ELECTRONIC INFORMATIN USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK, the specification of which,

(check one)

1. ☐ is attached hereto.

2. ☒ was filed on March 13, 2000 as
U.S. Application Serial No. 09/524,095
and was amended on ______________.

3. ☐ was filed on ______________ as
International PCT Application Serial No. ______________
and was amended on ______________.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, CFR § 1.56.

I hereby claim foreign priority benefits under Title 35, United States code, § 119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed:

| Prior Foreign Application(s) | | | Priority Benefits Claimed? |
|---|---|---|---|
| (Appl. No.) | (Country) | (Filing Date) | ☐Yes ☐No |
| (Appl. No.) | (Country) | (Filing Date) | ☐Yes ☐No |
| (Appl. No.) | (Country) | (Filing Date) | ☐Yes ☐No |

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below:

| | |
|---|---|
| (Application Serial No.) | (Filing Date) |
| (Application Serial No.) | (Filing Date) |

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s), or § 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

**Prior U.S. Application(s)**

(Application Serial No.) (Filing Date) (Status - patented, pending, abandoned)

(Application Serial No.) (Filing Date) (Status - patented, pending, abandoned)

And I hereby appoint the law firm of Hickman Stephens Coleman & Hughes, including Paul L. Hickman (Reg. No. 28,516); L. Keith Stephens (Reg. No. 32,632); Brian R. Coleman (Reg. No. 39,145); Michael J. Hughes (Reg. No. 29,077); Michael E. Melton (Reg. No. 32,276); Raymond E. Roberts (Reg. No. 38,597); Vidya R. Bhakar (Reg. No. 42,323); Larry B. Guernsey (Reg. No. 40,008); Douglas E. Mackenzie (Reg. No. 38,955); Michael D. Plimier (Reg. No. 43,004); Ronald B. Feece (Reg. No. P46,327); Stefanie M. Howell (Reg. No. P45,929); and Robert D. Hayden (Reg. No. 42,645) as my principal attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Send Correspondence To:

**HICKMAN STEPHENS COLEMAN & HUGHES, LLP**
**P.O. BOX 52037**
**Palo Alto, California 94303-0746**

Direct Telephone Calls To: **Raymond E. Roberts at telephone number (408) 558-9950**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Typewritten Full Name of Sole or First Inventor: Christine Halverson
Citizenship: USA
Inventor's signature: Christine Halverson
Date of Signature: 6·16·00
Residence: (City) San Jose
(State/Country) California/USA
Post Office Address: 1623 Fairorchard Avenue, San Jose, California 95125

Full Name of Second Joint Inventor (if any): Luc Julia
Citizenship: ~~USA~~ FRANCE
Inventor's signature:
Date of Signature: 6.21.00
Residence: (City) Menlo Park
(State/Country) California/USA
Post Office Address: 607 Menlo Avenue, Menlo Park, California 94025

Full Name of Third Joint Inventor (if any): Dimitris Voutsas
Citizenship: Greece
Inventor's signature:
Date of Signature: 6/16/00
Residence: (City) Thessaloniki
(State/Country) Greece
Post Office Address: 14 M. Pyrza Street, Neoi Epivates, Thessaloniki 57019, Greece

**Full Name of Fourth Joint Inventor (if any):** Adam Cheyer

**Citizenship:** USA

**Inventor's signature:**

**Date of Signature:** 6/22/00

**Residence:** **(City)** Palo Alto

**(State/Country)** California /USA

**Post Office Address:** 757 Cereza Drive, Palo Alto, California 94306

SECTORS
#3

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

| APPLICATION NUMBER | FILING/RECEIPT DATE | FIRST NAMED APPLICANT | ATTORNEY DOCKET NUMBER |
|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B |

Kevin J Zilka
P O Box 721030
San Jose, CA 95172-1030

**FORMALITIES LETTER**

*OC000000005370740*

Date Mailed: 09/01/2000

## NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

### FILED UNDER 37 CFR 1.53(b)

***Filing Date Granted***

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given TWO MONTHS from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is missing.
  *A properly signed oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date, is required.*
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of $65 for a small entity in compliance with 37 CFR 1.27, must be submitted with the missing items identified in this letter.

- **The balance due by applicant is $ 65.**

---

***A copy of this notice MUST be returned with the reply.***

M. Middleton
Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 2 - COPY TO BE RETURNED WITH RESPONSE

11/03/2000 HNGOR1 00000067 09608872

01 FC:205 65.00 OP

file://C:\APPS\PreExam\correspondence\2_B.xml 8/31/00

07-03-00

A

# IN THE UNIT STATES PATENT AND TRADEMA OFFICE

#4/A
LSS
4-2001
entered

CERTIFICATE OF EXPRESS MAILING
I hereby certify that this paper and the documents and/or fees referred to as attached therein are being deposited with the United States Postal Service on June 30, 2000 in an envelope as "Express Mail Post Office to Addressee" service under 37 CFR §1.10, Mailing Label Number **EK858788212US**, addressed to the Assistant Commissioner for Patents, Washington, DC 20231.

Kevin J. Zilka

Attorney Docket No.: SRI1P037B

First Named Inventor:

HALVERSEN, Christine

## UTILITY PATENT APPLICATION TRANSMITTAL (37 CFR. § 1.53(b))
(Continuation, Divisional or Continuation-in-part application)

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

☐ Duplicate for fee processing

Sir: This is a request for filing a patent application under 37 CFR. § 1.53(b) in the name of inventors: Christine Halversen, Luc Julia, Dimitris Voutsas, Adam Cheyer

For: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

This application is a ☒ Continuation ☐ Divisional ☐ Continuation-in-part

of prior Application No.: **09/524,095**, from which priority under 35 U.S.C. §120 is claimed.

Application Elements:

- ☒ 33 Pages of Specification, Claims and Abstract
- ☒ 07 Sheets of Drawings
- ☐ Declaration
  - ☐ Newly executed (original or copy)
  - ☐ Copy from a prior application (37 CFR 1.63(d) for a continuation or divisional). The entire disclosure of the prior application from which a copy of the declaration is herein supplied is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
    - ☐ Deletion of inventors Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).

Accompanying Application Parts:

- ☐ Assignment and Assignment Recordation Cover Sheet (recording fee of $40.00 enclosed)
- ☐ Power of Attorney
- ☐ 37 CFR 3.73(b) Statement by Assignee

(Revised 12/97, Pat App Trans 53(b) ContDivCIP)

☐ Information Disclos Statement with Form PTO-1449 ☐ pies of IDS Citations
☒ Preliminary Amendment
☒ Return Receipt Postcard
☐ Small Entity Statement(s) ☒ Statement filed in prior application. Status still proper and desired.
☐ Other:

Claim For Foreign Priority

☐ Priority of ________ Application No. _________ filed on __________ is claimed under 35 U.S.C. § 119.
☐ The certified copy has been filed in prior application U.S. Application No. _______
☐ The certified copy will follow.

Extension of Time for Prior Pending Application

☐ A Petition for Extension of Time is being concurrently filed in the prior pending application. A copy of the Petition for Extension of Time is attached.

Amendments

☐ Amend the specification by inserting before the first line the sentence: "This is a
☐ Continuation ☐ Continuation-in-part ☐ Divisional
application of copending prior
☐ Application No._________ filed on _____________,
☐ International Application _________ filed on _________ which designated the United States,
the disclosure of which is incorporated herein by reference."

☒ Cancel in this application original claims 2-55 of the prior application before calculating the filing fee. (*At least one original independent claim must be retained.*)

Fee Calculation (37 CFR § 1.16)

| | (Col. 1)<br>NO. FILED | (Col. 2)<br>NO. EXTRA | SMALL ENTITY<br>RATE | <br>FEE | OR | LARGE ENTITY<br>RATE | <br>FEE |
|---|---|---|---|---|---|---|---|
| BASIC FEE | | | $345 | $ 345 | OR | $690 | $ |
| TOTAL CLAIMS | 27 -20 = | 7 | x09 = | $ 63 | OR | x18 = | $ |
| INDEP CLAIMS | 3 -03 = | 0 | x39 = | $ | OR | x78 = | $ |
| [ ] Multiple Dependent Claim Presented | | | $130 = | $ | OR | $260 = | $ |
| * If the difference in Col. 1 is less than zero, enter "0" in Col. 2. | | | Total | $ 408 | OR | Total | $ |

☒ Check No. 137 in the amount of $ 408.00 is enclosed.

☒ The Commissioner is autho to charge any fees beyond the amount sed which may be required, or to credit any overpayment, to Deposit Account No. 50-1351 (Order No. SRI1P037B).

General Authorization for Petition for Extension of Time (37 CFR §1.136)

☒ Applicants hereby make and generally authorize any Petitions for Extensions of Time as may be needed for any subsequent filings. The Commissioner is also authorized to charge any extension fees under 37 CFR §1.17 as may be needed to Deposit Account No. 50-1351 (Order No. SRI1P037B).

☒ Please send correspondence to the following address:

**Kevin J. Zilka**
**P.O. BOX 721030**
**San Jose, California 95172-1030**

**Direct Telephone Calls To:** **Kevin J. Zilka at telephone number (408) 505-5100**

Date: June 30, 2000

**Kevin J. Zilka**
Registration No. 41,429

**PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

#5/B
LDJ
4-20-01
entered

In re the application of

Christine HALVERSEN et al.

Application No. 09/~~524,095~~ 608,872

Filed: March 13, 2000

For: NAVIGATING NETWORK BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

Docket:
SRI1P037B

Date: June 30, 2000

# Preliminary Amendment

Assistant Commissioner for Patents
and Trademarks
Washington, DC 20231

Dear Sir:

In regard to the above-named patent application, please enter the following amendments.

IN THE TITLE: ✓

Please delete "NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK", and insert therefore, --MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT--.

IN THE ABSTRACT: ✓

Please delete the Abstract and insert therefore --A system, method, and article of manufacture are provided for navigating an electronic data source by means of spoken language where a portion of the data link between a mobile information appliance of the user and the data

81

27

B

source utilizes wireless communication. When a spoken input request is received from a user who is using the mobile information appliance, it is interpreted. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is transmitted to the mobile information appliance.--

B1
end

IN THE SPECIFICATION:

On page 1, line 5, please delete "This is" and insert therefore, --This application is a continuation of an application entitled NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK which was filed on March 13, 2000 under serial number 09/524,095 and which is--.

B2

Please delete page 3, lines 3 to 32, and insert therefore, --The present invention addresses the above needs by providing a system, method, and article of manufacture for mobile navigation of network-based electronic data sources in response to spoken input requests. When a spoken input request is received from a user using a mobile information appliance that communicates with a network server via an at least partially wireless communications system, it is interpreted, such as by using a speech recognition engine to extract speech data from acoustic voice signals, and using a language parser to linguistically parse the speech data. The interpretation of the spoken request can be performed on a computing device locally with the user, such as the mobile information appliance, or remotely from the user. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is then transmitted to a client device of the user. If the network data source is a database, the navigation query is constructed in the format of a database query language.

B3

Typically, errors or ambiguities emerge in the interpretation of the spoken request, such that the system cannot instantiate a complete, valid navigational template. This is to be expected occasionally, and one preferred aspect of the invention is the ability to handle such errors and ambiguities in relatively graceful and user-friendly manner. Instead of simply rejecting such input and defaulting to traditional input modes or simply asking the user to try again, a preferred embodiment of the present invention seeks to converge rapidly toward instantiation of a valid navigational template by soliciting additional clarification from the user as necessary, either before or after a navigation of the data source, via multimodal input, i.e., by means of menu

28

B

selection or other input modalities including and in addition to spoken input. This clarifying, multi-modal dialogue takes advantage of whatever partial navigational information has been gleaned from the initial interpretation of the user's spoken request. This clarification process continues until the system converges toward an adequately instantiated navigational template, which is in turn used to navigate the network-based data and retrieve the user's desired information. The retrieved information is transmitted across the network and presented to the user on a suitable client display device.

B3 end

IN THE CLAIMS:

11

56 82

Please delete claims 1-55, and insert therefore the following claims ~~1-27~~:

B4

Sub C1

56 ~~1.~~ ~~(New)~~ A method for speech-based navigation of an electron~~ic data source located at one~~ or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication, comprising the steps of:

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d) utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network ~~server to the mobile information appliance of the user.~~

2 ~~57~~ ~~2.~~ (New) The method of claim ~~1~~ 1 ~~56~~, wherein the step of rendering the interpretation of the spoken request is performed at the one or more network servers.

3 ~~58~~ ~~3.~~ (New) The method of claim ~~1~~ 1 ~~56~~, wherein the step of rendering the interpretation of the spoken request is performed by the mobile information appliance.

4 ~~59~~ ~~4.~~ (New) The method of claim ~~1~~ 1 ~~56~~, further comprising the steps of soliciting additional input from the user, including user interaction in a modality different than the original request;

SRI1P037B

29

B

refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source.

5 ~~60~~ ~~5~~. (New) The method of claim ~~1~~ ~~56~~ 1, wherein the data link includes a cellular telephone system.

6 ~~61~~ ~~6~~. (New) The method of claim ~~1~~ ~~56~~ 1, wherein steps (a)-(d) are performed with respect to multiple users.

7 ~~62~~ ~~7~~. (New) The method of claim ~~1~~ ~~56~~ 1, wherein the mobile information appliance is a wireless telephone.

8 ~~63~~ ~~8~~. (New) The method of claim ~~1~~ ~~56~~ 1, wherein the mobile information appliance is a portable computing device.

B4

9 ~~64~~ ~~9~~. (New) The method of claim ~~8~~ ~~63~~ 8, wherein the portable computing device is a personal digital assistant.

SWB C2 ~~65~~ ~~10~~. ~~(New) A computer program embodied on a~~ computer ~~readable medium for~~ speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

30

B

11 ~~66~~ ~~11.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, wherein the rendering of the interpretation of the spoken request is performed at the one or more network servers.

12 ~~67~~ ~~12.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, wherein the rendering of the interpretation of the spoken request is performed by the mobile information appliance.

Rule 1.126

13 ~~68~~ ~~13.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, further comprising a code segment that solicits additional input from the user, including user interaction in a modality different than the original request; a code segment that refines the navigation query, based upon the additional input; and a code segment that uses the refined navigation query to select a portion of the electronic data source.

B4

14 ~~69~~ ~~14.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, wherein the data link includes a wireless telephone system.

15 ~~70~~ ~~15.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, wherein code segments (a)-(d) are executed with respect to multiple users.

16 ~~71~~ ~~16.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, wherein the mobile information appliance is a wireless telephone.

17 ~~72~~ ~~17.~~ (New) The computer program of claim 10 ~~65~~ ~~10~~, wherein the mobile information appliance is a portable computing device.

18 ~~73~~ ~~18.~~ (New) The computer program of claim 17 ~~72~~ ~~17~~, wherein the portable computing device is a personal digital assistant.

~~74~~ ~~19.~~ ~~(New) A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:~~

~~(a) a mobile information appliance operable to receive a spoken request for desired information from the user;~~

~~(b) spoken language processing logic, operable to render an interpretation of the spoken request;~~

31

B

Amd 03

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of a data link of the electronic communications infrastructure between a mobile information appliance of the user and the one or more network servers utilizes wireless communication.

B4

20 ~~75~~ ~~20~~. (New) The system of claim 19 ~~74~~ ~~19~~, wherein the spoken language processing logic renders the interpretation of the spoken request at the one or more network servers.

21 ~~76~~ ~~21~~. (New) The system of claim 19 ~~74~~ ~~19~~, wherein the spoken language processing logic renders the interpretation of the spoken request at the mobile information appliance.

22 ~~77~~ ~~22~~. (New) The system of claim 19 ~~74~~ ~~19~~, further comprising user interaction logic operable to solicit additional input from the user, including user interaction in a modality different than the original request; and query refining logic operable to refine the navigation query based upon the additional input; wherein the navigation logic users the refined navigation query to select a portion of the electronic data source.

23 ~~78~~ ~~23~~. (New) The system of claim 19 ~~74~~ ~~19~~, wherein the data link includes a cellular telephone system.

24 ~~79~~ ~~24~~. (New) The system of claim 19 ~~74~~ ~~19~~, wherein the system operates with respect to multiple users.

25 ~~80~~ ~~25~~. (New) The system of claim 19 ~~74~~ ~~19~~, wherein the mobile information appliance is a wireless telephone.

26 ~~81~~ ~~26~~. (New) The system of claim 19 ~~74~~ ~~19~~, wherein the mobile information appliance is a portable computing device.

32

B

Rule 126

27 ~~822~~ ~~27~~. (New) The system of claim 26 ~~81~~ ~~26~~, wherein the portable computing device is a personal digital assistant.

B4 end

---

In the event a telephone conversation would expedite the prosecution of this application, the Examiner may reach the undersigned at (408) 505-5100. If any fees are due in connection with the filing of this paper, then the Commissioner is authorized to charge such fees to Deposit Account No. 50-1351 (Order No. SRI1P037B). A duplicate copy of the transmittal is enclosed for this purpose.

Respectfully submitted,

Kevin J. Zilka
Registration No. 41,429

P.O. Box 721030
San Jose, CA 95172
Telephone: (408) 505-5100

33

B

6

**UNITED STATES DEPARTMENT OF COMMERCE**
**United States Patent and Trademark Office**
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | | ATTORNEY DOCKET NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/00 | HALVERSEN | C | SRILP037B |

024277 TM02/0424
Kevin J. Zilka
PO Box 721030
San Jose CA 95172

| EXAMINER | |
|---|---|
| BACKER,F | |
| **ART UNIT** | **PAPER NUMBER** |
| 2155 | 6 |

**DATE MAILED:** 04/24/01

**Please find below and/or attached an Office communication concerning this application or proceeding.**

**Commissioner of Patents and Trademarks**

PTO-90C (Rev.11/00)

1- File Copy

| **Office Action Summary** | **Application No.**<br>09/608,872 | **Applicant(s)**<br>HALVERSEN ET AL. |
|---|---|---|
| | **Examiner**<br>Firmin Backer | **Art Unit**<br>2155 |

***-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --***

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136 (a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on _30 June 2000_.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) _56-82_ is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _56-82_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claims _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are objected to by the Examiner.

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved.

12)☐ The oath or declaration is objected to by the Examiner.

**Priority under 35 U.S.C. § 119**

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

**Attachment(s)**

15) ☒ Notice of References Cited (PTO-892)
16) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
17) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .
18) ☐ Interview Summary (PTO-413) Paper No(s). _____ .
19) ☐ Notice of Informal Patent Application (PTO-152)
20) ☐ Other: .

## DETAILED ACTION

This is in response to a letter for patent filed on June 30th, 2000 in which claims 56-82 are presented for examination. Claims 56-82 are pending in the letter.

### *Double Patenting*

1. A rejection based on double patenting of the "same invention" type finds its support in the language of 35 U.S.C. 101 which states that "whoever invents or discovers any new and useful process ... may obtain a patent therefor ..." (Emphasis added). Thus, the term "same invention," in this context, means an invention drawn to identical subject matter. See *Miller v. Eagle Mfg. Co.*, 151 U.S. 186 (1894); *In re Ockert*, 245 F.2d 467, 114 USPQ 330 (CCPA 1957); and *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970).

A statutory type (35 U.S.C. 101) double patenting rejection can be overcome by canceling or amending the conflicting claims so they are no longer coextensive in scope. The filing of a terminal disclaimer cannot overcome a double patenting rejection based upon 35 U.S.C. 101.

2. Claims 56-82 are provisionally rejected under 35 U.S.C. 101 as claiming the same invention as that of claims 56-126 of copending Application No. 09/524,095. Although the conflicting claims are not identical, they are not patentably distinct. It would have been obvious to one of ordinary skill in the art to observed that the omission of the limitations "***soliciting additional input from the user, including user interaction in a modality different that the original request and, refining the navigation query, based upon the additional input***", of applicant claims 56-82 are already in the Co-pending application 09/524,095, as such they are obvious variation of the inventive concept defined in claims 56-126 of the Co-pending application 09/524,095. See In re Karlson, 136USPQ 184 (CCPA 1963). This is a provisional double patenting rejection since the conflicting claims have not in fact been patented.

### *Claim Rejections - 35 USC § 102*

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent.

4. Claims 56-82 are rejected under 35 U.S.C. 102(e) as being anticipated by Levin et al. (U.S. Patent No. 6,173,279).

5. As per claim 56, Levin et al teach a method for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising receiving a spoken request (*receive a natural language query*) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user; rendering an interpretation (*creating a semantic representation*) of the spoken request, constructing a navigation (*generating search)* query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)

6. As per claim 57, 58, 62-64, Levin et al teach a method of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

7. As per claim 59, Levin et al teach a method of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

8. As per claim 60, Levin et al teach a method wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

9. As per claim 61, Levin et al teach a method wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

10. As per claim 65, Levin et al teach a computer system for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising a code segment receiving a

spoken request (*receive a natural language query*) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user; a code segment rendering an interpretation (*creating a semantic representation*) of the spoken request, a code segment constructing a navigation (*generating search)* query based upon the interpretation; a code segment utilizing the navigation query to select a portion of the electronic data source; and a code segment transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)

11. As per claim 66, 67, 71-73, Levin et al teach a system of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

12. As per claim 68, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

13. As per claim 69, Levin et al teach a system wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

14. As per claim 70, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

15. As per claim 74, Levin et al teach a system for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising receiving a spoken request (*receive a natural language query*) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user; rendering an interpretation (*creating a semantic representation*) of the spoken request, constructing a navigation (*generating search)* query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)

16. As per claim 75, 76, 80-81, Levin et al teach a method of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

17. As per claim 77, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the

navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

18. As per claim 78, Levin et al teach a system wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

19. As per claim 79, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

## *Conclusion*

20. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. (6,192,338).

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Firmin Backer whose telephone number is 703-305-0624. The examiner can normally be reached on Mon-Thu 8:30-6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Sheikh Ayaz can be reached on 703-305-9648. The fax phone numbers for the organization where this application or proceeding is assigned are 703-305-3718 for regular communications and 703-305-5352 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

Firmin Backer
April 9, 2001

AYAZ SHEIKH
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

PAGE 1 OF 1

| | | | | |
|---|---|---|---|---|
| FORM PTO-892<br>NOTICE OF REFERENCES CITED | U.S. DEPARTMENT OF COMMERCE<br>PATENT AND TRADEMARK OFFICE | SERIAL NO.<br>09/608,872 | GROUP ART UNIT 2155<br>~~2781~~ | ATTACHMENT TO PAPER NO. 10 |
| | | APPLICANT(S)<br>HALVERSEN ET AL. | | |

**U.S. PATENT DOCUMENTS**

| * | | DOCUMENT NO. | DATE | NAME | CLASS | SUB-CLASS | FILING DATE |
|---|---|---|---|---|---|---|---|
| | A | 6,192,338 | 2/2001 | Haszto et al | 704 | 257 | |
| | B | 6,173,279 | 1/2001 | Levin et al. | 707 | 5 | |
| | C | | | | | | |
| | D | | | | | | |
| | E | | | | | | |
| | F | | | | | | |
| | G | | | | | | |
| | H | | | | | | |
| | I | | | | | | |
| | J | | | | | | |
| | K | | | | | | |

**FOREIGN PATENT DOCUMENTS**

| * | | DOCUMENT NO. | DATE | COUNTRY | NAME | CLASS | SUB-CLASS |
|---|---|---|---|---|---|---|---|
| | L | | | | | | |
| | M | | | | | | |
| | N | | | | | | |
| | O | | | | | | |
| | P | | | | | | |
| | Q | | | | | | |

**OTHER REFERENCES (Including Author, Title, Date, Pertinent Pages, Etc.)**

| | |
|---|---|
| R | |
| S | |
| T | |
| U | |

| EXAMINER | DATE | |
|---|---|---|
| Firmin Backer | April 9, 2001 | Form892ccs2106b |

* A copy of this reference is not being furnished with this office action.
(See Manual of Patent Examining Procedure, section 707.05(a).)

OIPE JC170
APR 3 0 2001
PATENT & TRADEMARK OFFICE

2155

**PATENT**

#7
LDD
5-8-01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the application of:

Halverson et al.

Application No. 09/608,872

Filed: 06/30/2000

For: MOBILE NAVIGATION OF NETWORK -BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

Group Art Unit: 2741

Examiner: Unassigned

Atty. Docket No. SRI1P037B/
44454/03450

Date: April 27, 2001

RECEIVED
MAY 4 - 2001
Technology Center 2100

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, DC 20231 on April 27, 2001.

Signed: Erica L. Mann
Erica L. Mann

## INFORMATION DISCLOSURE STATEMENT UNDER 37 CFR §§ 1.56 AND 1.97(c)

Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

The references listed in the attached PTO Form 1449, copies of which are attached, may be material to examination of the above-identified patent application. Applicants submit these references in compliance with their duty of disclosure pursuant to 37 CFR §§ 1.56 and 1.97. The Examiner is requested to make these references of official record in this application.

This Information Disclosure Statement is not to be construed as a representation that a search has been made, that additional information material to the examination of this application does not exist, or that these references indeed constitute prior art.

This Information Disclosure Statement is believed to be filed before the mailing date of a first Office Action on the merits. Accordingly, it is believed that no fees are due in connection with the filing of this Information Disclosure Statement. However, if it is determined that any fees are due, the Commissioner is hereby authorized to charge such fees to Deposit Account 03-0683 (Order No. 44454/03450/SRI1P037B).

Respectfully submitted,
CARLTON FIELDS

Dominic M. Kotab
Reg. No. 42,762

P.O. Box 721030
San Jose, CA 95172-1030
Telephone: (408) 271-2300

#7

| | |
|---|---|
| **Form 1449 (Modified)**<br><br>**Information Disclosure Statement By Applicant**<br><br>(Use Several Sheets if Necessary) | Atty. Docket No.<br>SRI1P037B<br>Applicant:<br>Halverson et al.<br>Filing Date:<br>06/30/2000 |
| | Application No.:<br>09/608,872<br><br><br>Group Art Unit:<br>~~2741~~ 2155 |

**U.S. Patent Documents**

| Examiner Initial | No. | Patent No. | Date | Patentee | Class | Sub-class | Filing Date |
|---|---|---|---|---|---|---|---|
| FO | A | 6,026,388 | 02/15/00 | Liddy et al. | 707 | 1 | 08/14/96 |
| | B | 6,102,030 | 01/04/00 | French- St. George et al. | 704 | 275 | 04/21/98 |
| | C | 6,003,072 | 12/14/99 | Gerritsen et al. | 709 | 218 | 06/30/94 |
| | D | 5,890,123 | 03/30/99 | Brown et al. | 704 | 275 | 06/05/95 |
| | E | 5,855,002 | 12/29/98 | Armstrong | 704 | 270 | 06/11/96 |
| | F | 5,963,940 | 10/05/99 | Liddy et al. | 707 | 5 | 08/14/96 |
| | G | 5,805,775 | 09/08/98 | Eberman et al. | 395 | 12 | 02/02/96 |
| | H | 5,802,526 | 09/01/98 | Fawcett et al. | 707 | 10 | 04/18/96 |
| | I | 5,794,050 | 08/11/98 | Dahlgren et al. | 395 | 708 | 10/02/97 |
| | J | 5,774,859 | 06/30/98 | Houser et al. | 704 | 275 | 01/03/95 |
| FO | K | 5,748,974 | 05/05/98 | Johnson | 395 | 759 | 12/13/94 |

**Foreign Patent or Published Foreign Patent Application**

| Examiner Initial | No. | Document No. | Publication Date | Country or Patent Office | Class | Sub-class | Translation | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Yes | No |
| | L | | | | | | | |
| | M | | | | | | | |
| | N | | | | | | | |
| | O | | | | | | | |
| | P | | | | | | | |

**Other Documents**

| Examiner Initial | No. | Author, Title, Date, Place (e.g. Journal) of Publication |
|---|---|---|
| FO | R | Stent, Amanda et al., "The CommandTalk Spoken Dialogue System", SRI International |
| FO | S | Moore, Robert et al., "CommandTalk: A Spoken-Language Interface for Battlefield Simulations", October 23, 1997, SRI International |
| FO | T | Dowding, John et al., "Interpreting Language in Context in CommandTalk", February 5, 1999, SRI International |

| | |
|---|---|
| Examiner | Date Considered 9/27/02 |

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

#7

OIPE
APR 3 0 2001
PATENT & TRADEMARK OFFICE

| | |
|---|---|
| **Form 1449 (Modified)**<br>**Information Disclosure Statement By Applicant**<br>(Use Several Sheets if Necessary) | Atty. Docket No. SRI1P037B<br>Applicant: Halverson et al.<br>Filing Date: 06/30/2000 |
| | Application No.: 09/608,872<br>Group Art Unit: ~~2741~~ 2155 |

RECEIVED
MAY 4 2001
Technology Center 2100

**U.S. Patent Documents**

| Examiner Initial | No. | Patent No. | Date | Patentee | Class | Sub-class | Filing Date |
|---|---|---|---|---|---|---|---|
| FO | A | 5,729,659 | 03/17/98 | Potter | 395 | 2.79 | 06/06/95 |
| | B | 5,721,938 | 02/24/98 | Stuckey | 395 | 754 | 06/07/95 |
| | C | 5,608,624 | 03/04/97 | Luciw | 395 | 794 | 05/15/95 |
| | D | 5,519,608 | 05/21/96 | Kupiec | 364 | 419.08 | 06/24/93 |
| | E | 5,434,777 | 07/18/95 | Luciw | 364 | 419.13 | 03/18/94 |
| | F | 5,386,556 | 01/31/95 | Hedin et al. | 395 | 600 | 10/23/92 |
| FO | G | 5,197,005 | 03/23/93 | Shwartz et al. | 364 | 419 | 05/01/89 |
| | H | | | | | | |
| | I | | | | | | |
| | J | | | | | | |
| | K | | | | | | |

**Foreign Patent or Published Foreign Patent Application**

| Examiner Initial | No. | Document No. | Publication Date | Country or Patent Office | Class | Sub-class | Translation Yes | Translation No |
|---|---|---|---|---|---|---|---|---|
| | L | | | | | | | |
| | M | | | | | | | |
| | N | | | | | | | |
| | O | | | | | | | |
| | P | | | | | | | |

**Other Documents**

| Examiner Initial | No. | Author, Title, Date, Place (e.g. Journal) of Publication |
|---|---|---|
| FO | R | http://www.ai.sri.com/~oaa/infowiz.html, "InfoWiz: An Animated Voice Interactive Information System, May 8, 2000 |
| FO. | S | Dowding, John, "Interleaving Syntax and Semantics in an Efficient Bottom-up Parser", SRI International |
| FO | T | Moore, Robert et al., "Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS", SRI International |

| | |
|---|---|
| Examiner | Date Considered 9/27/02 |

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

Pg. 2 of 3

#7

| | | |
|---|---|---|
| **Form 1449 (Modified)**<br>**Information Disclosure Statement By Applicant**<br>(Use Several Sheets if Necessary) | Atty. Docket No.<br>SRI1P037B<br>Applicant:<br>Halverson et al.<br>Filing Date:<br>06/30/2000 | Application No.:<br>09/608,872<br><br><br>Group Art Unit:<br>~~2741~~ 2155 |

**U.S. Patent Documents**

| Examiner Initial | No. | Patent No. | Date | Patentee | Class | Sub-class | Filing Date |
|---|---|---|---|---|---|---|---|
| | A | | | | | | |
| | B | | | | | | |
| | C | | | | | | |
| | D | | | | | | |
| | E | | | | | | |
| | F | | | | | | |
| | G | | | | | | |
| | H | | | | | | |
| | I | | | | | | |
| | J | | | | | | |
| | K | | | | | | |

**Foreign Patent or Published Foreign Patent Application**

| Examiner Initial | No. | Document No. | Publication Date | Country or Patent Office | Class | Sub-class | Translation Yes | Translation No |
|---|---|---|---|---|---|---|---|---|
| | L | | | | | | | |
| | M | | | | | | | |
| | N | | | | | | | |
| | O | | | | | | | |
| | P | | | | | | | |

**Other Documents**

| Examiner Initial | No. | Author, Title, Date, Place (e.g. Journal) of Publication |
|---|---|---|
| | R | Dowding, John et al., "Gemini: A Natural Language System For Spoken-Language Understanding", SRI International |
| | S | |
| | T | |

| | |
|---|---|
| Examiner | Date Considered 9/27/02 |

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

OIPE JC157
MAY 14 2001
PATENT & TRADEMARK OFFICE

4B17

GAU-2155
#8

Attorney Docket No.: SRI1P037B

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: HALVERSON, CHRISTINE
SERIAL NO.: 09/608,872
FILED: 6/30/00
TITLE: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATIONUSING SPOKEN INPUT

RECEIVED
MAY 17 2001
Technology Center 2100

---

## ASSOCIATE POWER OF ATTORNEY

Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

I hereby appoint: C. Douglas McDonald (Reg. No. 26,659)

whose post office address is

Carlton Fields, P.A.
P. O. Box 3239
Tampa, Florida 33601-3239

as my associate attorney in the above-entitled application, to prosecute this application, to make alterations and amendments therein, and to transact all business in the Patent and Trademark Office connected therewith.

Please continue to address all future communications to:

Carlton Fields, LLP
P. O. Box 721030
San Jose, CA 95172-1030

Respectfully submitted

Date: MAY 7, 2001

Kevin J. Zilka (Reg. No. 41,429)
Dominic Kotab (Reg. No. 42,762)
Carlton Fields LLP
P.O. Box 721030
San Jose, CA 95172-1030
Telephone: (408) 271-2300
Fax: (408) 275-9579

TPA#1680358.01

#9

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.: 09/608,872
INVENTOR: Halverson, Christine
TITLE: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATIONUSING SPOKEN INPUT

FILING DATE: 6/30/00
ATTORNEY DOCKET NO. SRI1P037B

## NOTICE OF CHANGE OF CORRESPONDENCE ADDRESS

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

Please change the correspondence address relating to the above-identified application as follows:

C. Douglas McDonald , Esq.
Carlton Fields, et al.
P.O. Box 3239
Tampa, FL 33601-3239

Respectfully submitted,

C. Douglas McDonald

Date: May 10, 2001

C. Douglas McDonald
Reg. No. 26,659
CARLTON FIELDS, P.A.
P.O. Box 3239
Tampa, FL 33601-3239
(813) 223-7000
Attorney of Record

TPA#1524975.01

U.S. Patent and Trademark Office; U.S. [illegible] MENT OF COMMERCE
Under the Paperwork Reduction Act of 1[illegible] no persons are required to respond to a collection of information unless it [illegible] valid OMB control number.

## PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)

Docket Number (Optional)
SRI 1P037B

#10
LDJ
10-01-0

OIPE JC95
SEP 2 1 2001
PATENT & TRADEMARK OFFICE

In re Application of HALVERSON, et al

Application Number 09/608,872
Filed June 30, 2000

For Mobile Navigation of Network-Based Electronic Information Using Spoken Input

Group Art Unit 2155
Examiner F. Backer

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a response in the above identified application.

The requested extension and appropriate non-small-entity fee are as follows (check time period desired):

| | |
|---|---|
| ☐ One month (37 CFR 1.17(a)(1)) | $ |
| ☒ Two months (37 CFR 1.17(a)(2)) | $390.00 |
| ☐ Three months (37 CFR 1.17(a)(3)) | $ |
| ☐ Four months (37 CFR 1.17(a)(4)) | $ |
| ☐ Five months (37 CFR 1.17(a)(5)) | $ |

☒ Applicant claims small entity status. See 37 CFR 1.27. Therefore, the fee amount shown above is reduced by one-half, and the resulting fee is: $ 195.00 .

☒ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Commissioner has already been authorized to charge fees in this application to a Deposit Account.

☒ The Commissioner is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 20-0782 .
I have enclosed a duplicate copy of this sheet.

RECEIVED
SEP 2 6 2001
Technology Center 2100

I am the ☐ applicant/inventor.

☐ assignee of record of the entire interest. See 37 CFR 3.71
Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).

☒ attorney or agent of record.

☐ attorney or agent under 37 CFR 1.34(a).

Registration number if acting under 37 CFR 1.34(a). ______ .

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

September 19, 2001
Date

Signature

KIN-WAH TONG, Reg. No. 39,400
Typed or printed name

09/25/2001 MWOLDER1 00000030 09608872
01 FC:216 195.00 OP

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of _____ forms are submitted.

Burden Hour Statement: This form is estimated to take 0.1 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

SRI/4116-6

#11
LDD 10-01-01

OIPE
SEP 21 2001
PATENT & TRADEMARK OFFICE

SEP 21 2001
PATENT & TRADEMARK OFFICE

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

## PATENT APPLICATION

| | | | |
|---|---|---|---|
| Applicant(s): | **HALVERSON, et al.** | Atty. Docket No. | **SRI 1P037B** |
| Serial No.: | **09/608,872** | Group Art Unit: | **2155** |
| Filed: | **June 30, 2000** | Examiner: | **F. Backer** |
| Title: | **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT** | | |

RECEIVED
SEP 26 2001
Technology Center 2100

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

## REVOCATION OF PREVIOUS POWER OF ATTORNEY AND NEW APPOINTMENT

The undersigned assignee of the above-identified application hereby revokes all previous Powers of Attorney and appoints the following attorneys with full power to prosecute the application, to make alterations and amendments therein, and to transact all business in the United States Patent and Trademark Office connected therewith and with full power of substitution and revocation:

Raymond R. Moser, Jr.; Reg. No. 34,682; Kin-Wah Tong, Reg. No. 39,400; Robert Brush, Reg. No. 45,710; Steven Weiner, Reg. No. 38,360; and Edward E. Davis, Reg. No. 35,112.

## CHANGE OF CORRESPONDENCE ADDRESS

Please change the correspondence address for the above-identified application to:

Thomason, Moser & Patterson, LLP
595 Shrewsbury Avenue – Suite 100
Shrewsbury, New Jersey 07702

Please direct all telephone calls to: Kin-Wah Tong, telephone # (732) 530-9404

SRI/4116-6

## CERTIFICATE UNDER 37 C.F.R. § 3.73(B)

SRI International, a corporation of the State of California, certifies that it is the assignee of the entire right, title and interest in the patent application identified above by virtue of:

An Assignment from the inventor(s) of the parent patent application that is claimed as priority in the above-identified patent application. The Assignment was recorded in the United States Patent and Trademark Office, for which a copy thereof is attached.

The undersigned (whose title is supplied below) is empowered to act on behalf of the assignee.

Respectfully submitted,

Date: 9/11/01

~~EDWARD E. DAVIS, Assistant Secretary~~
STEVEN WEINER, VICE PRESIDENT

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Telephone No.: 650-859-3115

# ASSIGNMENT OF PATENT APPLICATION

(Not Accompanying Application)

Whereas I/we the undersigned inventor(s) have invented certain new and useful improvements as set forth in the patent application entitled:

## NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

for which I/we have executed an application for a United States Letters Patent which was filed in the U.S. Patent and Trademark Office on March 13, 2000, and which bears the Application No. 09/524,095.

For good and valuable consideration, the receipt and sufficiency of which is hereby acknowledged, I/we the undersigned inventor(s) hereby:

1) Sell(s), assign(s) and transfer(s) to **SRI International,** a California non-profit corporation having a place of business at 333 Ravenswood Avenue, Menlo Park, California 94025, (hereinafter referred to as "ASSIGNEE"), the entire right title and interest in any and all improvements and inventions disclosed in, application(s) based upon, and Patent(s) (including foreign patents) granted upon the information which is disclosed in the above referenced application.

2) Authorize and request the Commissioner of Patents to issue any and all Letters Patents resulting from said application or any division(s), continuation(s), substitutes(s) or reissue(s) thereof to the ASSIGNEE.

3) Agree to execute all papers and documents and, entirely at the ASSIGNEE's expense, perform any acts which are reasonably necessary in connection with the prosecution of said application, as well as any derivative and applications thereof, foreign applications based thereon, and/or the enforcement of patents resulting from such applications.

4) Agree that the terms, covenants and conditions of this assignment shall inure to the benefit of the Assignee, its successors, assigns and other legal representative, and shall be binding upon the inventor(s), as well as the inventor's heirs, legal representatives and assigns.

5) Warrant and represent that I/we have not entered, and will not enter into any assignment, contract, or understanding that conflicts with this assignment.

Signed on the date(s) indicated beside my (our) signature(s).

1) Signature: ____________ Date: 6-16-00.
Typed Name: Christine Halverson

2) Signature: ____________ Date: ________
Typed Name: Luc Julia

3) Signature: ____________ Date: 6/16/00
Typed Name: Dimitris Voutsas

4) Signature: ____________ Date: 6/22/00
Typed Name: Adam Cheyer

Attny Docket No. SRI1P037

# ASSIGNMENT OF PATENT APPLICATIO

(Not Accompanying Application)

Whereas I/we the undersigned inventor(s) have invented certain new and useful improvements as set forth in the patent application entitled:

**NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK**

for which I/we have executed an application for a United States Letters Patent which was filed in the U.S. Patent and Trademark Office on March 13, 2000, and which bears the Application No. 09/524,095.

For good and valuable consideration, the receipt and sufficiency of which is hereby acknowledged, I/we the undersigned inventor(s) hereby:

1) Sell(s), assign(s) and transfer(s) to **SRI International**, a California non-profit corporation having a place of business at 333 Ravenswood Avenue, Menlo Park, California 94025, (hereinafter referred to as "ASSIGNEE"), the entire right title and interest in any and all improvements and inventions disclosed in, application(s) based upon, and Patent(s) (including foreign patents) granted upon the information which is disclosed in the above referenced application.

2) Authorize and request the Commissioner of Patents to issue any and all Letters Patents resulting from said application or any division(s), continuation(s), substitutes(s) or reissue(s) thereof to the ASSIGNEE.

3) Agree to execute all papers and documents and, entirely at the ASSIGNEE's expense, perform any acts which are reasonably necessary in connection with the prosecution of said application, as well as any derivative and applications thereof, foreign applications based thereon, and/or the enforcement of patents resulting from such applications.

4) Agree that the terms, covenants and conditions of this assignment shall inure to the benefit of the Assignee, its successors, assigns and other legal representative, and shall be binding upon the inventor(s), as well as the inventor's heirs, legal representatives and assigns.

5) Warrant and represent that I/we have not entered, and will not enter into any assignment, contract, or understanding that conflicts with this assignment.

Signed on the date(s) indicated beside my (our) signature(s).

1) Signature: Christine Halverson Date: 6-16-00.
Typed Name: Christine Halverson

2) Signature: Date: 6·20.00
Typed Name: Luc Julia

3) Signature: Date: 6/16/00
Typed Name: Dimitris Voutsas

4) Signature: Date:
Typed Name: Adam Cheyer

Attny Docket No. SRI1P037

# ASSIGNMENT OF PATENT APPLICATION

(Not Accompanying Application)

Whereas I/we the undersigned inventor(s) have invented certain new and useful improvements as set forth in the patent application entitled:

**NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK**

for which I/we have executed an application for a United States Letters Patent which was filed in the U.S. Patent and Trademark Office on March 13, 2000, and which bears the Application No. 09/524,095.

For good and valuable consideration, the receipt and sufficiency of which is hereby acknowledged, I/we the undersigned inventor(s) hereby:

1) Sell(s), assign(s) and transfer(s) to **SRI International,** a California non-profit corporation having a place of business at 333 Ravenswood Avenue, Menlo Park, California 94025, (hereinafter referred to as "ASSIGNEE"), the entire right title and interest in any and all improvements and inventions disclosed in, application(s) based upon, and Patent(s) (including foreign patents) granted upon the information which is disclosed in the above referenced application.

2) Authorize and request the Commissioner of Patents to issue any and all Letters Patents resulting from said application or any division(s), continuation(s), substitutes(s) or reissue(s) thereof to the ASSIGNEE.

3) Agree to execute all papers and documents and, entirely at the ASSIGNEE's expense, perform any acts which are reasonably necessary in connection with the prosecution of said application, as well as any derivative and applications thereof, foreign applications based thereon, and/or the enforcement of patents resulting from such applications.

4) Agree that the terms, covenants and conditions of this assignment shall inure to the benefit of the Assignee, its successors, assigns and other legal representative, and shall be binding upon the inventor(s), as well as the inventor's heirs, legal representatives and assigns.

5) Warrant and represent that I/we have not entered, and will not enter into any assignment, contract, or understanding that conflicts with this assignment.

Signed on the date(s) indicated beside my (our) signature(s).

1) Signature: ____________ Date: 6-16-00.
Typed Name: Christine Halverson

2) Signature: ____________ Date: ________
Typed Name: Luc Julia

3) Signature: ____________ Date: 6/16/00
Typed Name: Dimitris Voutsas

4) Signature: ____________ Date: ________
Typed Name: Adam Cheyer

O I P E
SEP 2 1 2001
JC95
PATENT & TRADEMARK OFFICE

#12
LDD
10-01-01

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

## PATENT APPLICATION

RECEIVED
SEP 26 2001
Technology Center 2100

Applicant: **Halverson et al.**

Case: **SRI1P037B**

Serial No.: **09/608,872** Filed: **June 30, 2000**

Group Art Unit: **2155**

Examiner: **Firmin Backer**

Title: **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT**

ASSISTANT COMMISSIONER FOR PATENTS
Box Non-Fee Amendment
Washington, D. C. 20231

S I R:

## RESPONSE UNDER 37 C.F.R. § 1.111

This response addresses the Office Action dated April 24, 2001 (Paper No. 10).

## REMARKS

In view of the following discussion, the Applicants submit that none of the claims now pending in the application are anticipated under the provisions of 35 U.S.C. § 102. Thus, the Applicants believe that all of these claims are now in allowable form.

## I. REJECTION OF CLAIMS 56-82 UNDER DOUBLE PATENTING

The Examiner provisionally rejected claims 56-82 in Paragraphs 1-2 of the Office Action based on statutory type double patenting under 35 U.S.C. § 101 as claiming the same invention as that of claims 56-126 of copending Application No. 09/524,095. Applicants respectfully traverse the rejection.

First, the Examiner noted that "it would have been obvious to one of ordinary skill in the art to observe that the omission of the limitations '**soliciting additional input**

**from the user, including user interaction in a modality different tha[n] the original request and, refining the navigation query, based upon the additional input**'. After noting the differences between the scope of the claims between the two applications, the Examiner then concluded that claims 56-82 "are obvious variation of the inventive concept defined in claims 56-126 of co-pending application 09/524,095".

Applicants direct the Examiner's attention to the fact that there are two types of double patenting rejections: "statutory" and "non-statutory (obviousness-type)". MPEP 804 states that "[i]n determining whether a statutory basis for a double patenting rejection exists, the question to be asked is: Is the same invention being claimed twice?" "A reliable test for double patenting under 35 U.S.C. 101 is whether a claim in the application could be literally infringed without literally infringing a corresponding claim in the patent". Given the substantial differences between the claims of the two applications as noted by the Examiner, Applicants respectfully submit that applying the statutory double patenting test as promoted in the MPEP would not produce a statutory double patenting rejection in the present application. As such, Applicants submit that the present statutory double patenting rejection against claims 56-82 is inappropriate.

Second, it should be noted that the present application is a continuation of the co-pending application 09/524,095. As such, if and when these two applications mature into issued patents, both patents will have the same term. Thus, given the differences between the scope of the claims of both applications and the fact that both applications will expire at the same time (if issued), Applicants respectfully submit that statutory double patenting rejection against claims 56-82 is inappropriate.

## II. REJECTION OF CLAIMS 56-82 UNDER 35 U.S.C. § 102

The Examiner has rejected claims 56-82 in Paragraphs 4-19 of the Office Action as being anticipated by the Levin et al. patent (US Patent 6,173,279 issued January 9, 2001, hereinafter referred to as Levin). The rejection is respectfully traversed.

Levin teaches "a method of using at least one natural language query to retrieve information from one or more data resources and further performing a requested action using the retrieved information is disclosed". (See Levin, Column 2, lines 15-18)

Namely, Levin teaches a method for using natural language query to obtain information, where upon receipt of the requested information, a desired action is executed based upon the requested information. To illustrate, Levin provides the example, where a user employs natural language to request the telephone number of a restaurant. Upon receipt of the telephone number, the telephone number is actually dialed for the user. (See Levin, Column 3 line 62 to Column 4, line 1)

In contrast, Levin fails to teach or suggest the novel concept of speech-based navigation <u>where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where, in turn, the selected electronic data source from the network server is transmitted to the mobile information appliance of the user</u>. Specifically, Applicants' independent claims 56, 65 and 74 positively recite:

> 56. A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilize wireless communication, comprising the steps of:
>
> (a) <u>receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user;</u>
>
> (b) rendering an interpretation of the spoken request;
>
> (c) constructing a navigation query based upon the interpretation;
>
> (d)utilizing the navigation query to select a portion of the electronic data source; and
>
> (e) <u>transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user</u>. (emphasis added)
>
> 65. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication, comprising:
>
> (a) <u>a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user;</u>
>
> (b) a code segment that renders an interpretation of the spoken request.
>
> (c) a code segment that constructs a navigation query based upon the

interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

74. A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

(a) a mobile information appliance operable to receive a spoken request for desired information from the user;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of a data link of the electronic communications infrastructure between a mobile information appliance of the user and the one or more network servers utilizes wireless communication. (emphasis added)

Applicants' invention teaches a novel method and apparatus for speech-based navigation where the method receives spoken request for desired information from the user utilizing the **mobile information appliance** of the user and where, in turn, the selected electronic data source from the network server is **transmitted to the mobile information appliance** of the user. Specifically, Applicants address the criticality of providing speech-based navigation via a mobile, i.e., wireless communication, approach in addition to spoken natural language. It has been noted that with the proliferation of various mobile appliances, it would be advantageous to allow these mobile appliances to access the same vastness of electronic data sources that are available to hard-wired appliances like a desktop computer. However, the very essence of a mobile appliance is its portability, small size and ease of use. As such, unlike hard-wired appliances, mobile appliances are not equipped with large bulky input devices. In fact, even if the mobile appliance is equipped with extensive input devices, most users would still find these "shrunken" input devices to be cumbersome and difficult to use, e.g., an electronic representation of a keyboard on a PDA and the like.

To further exacerbate the problem, obtaining information from an electronic data source may require extensive and complex interaction between the user's mobile appliance and the system holding the electronic data source. Thus, the limited or cumbersome input/output capability of a mobile appliance presents a substantial barrier to its ability to access a data resource that requires extensive and complex interaction.

To address this criticality, Applicants disclose a speech-based navigation method that is deployed in conjunction with mobile appliances. To illustrate, the user can request via a mobile appliance, e.g., a cellular telephone, all the names of a particular ethnic restaurant on a particular street. Clearly, this request is rather complex given the limited input capability (generally a numeric keypad) of a cellular phone. Without additional input devices, this complex request may require numerous interactions between the user and a remote data resource, e.g., long repeated sequences of presenting a menu, scrolling within the menu and selecting the desired information within the menu and so on for the next menu and beyond. Such tedium discourages a user from attempting to acquire complex information via mobile appliances.

In contrast, Applicants' invention allows the complex request to be received as a spoken request directly via the user's mobile information appliance, thereby substantially reducing the amount of interaction of the user with the remote data resource. The present method will interpret and construct a navigation query that is utilized to obtain the selected data. For example, if the navigation query produces three possible results, then the results can be simply transmitted to the user via a menu on the screen of the mobile appliance.

In contrast, Levin teaches that "[u]sing a personal computer (PC) 102, a user establishes a connection with packet network 108 via an access server 106". Levin then states that "[t]he user may also use a telephone 103 to connect to the packet network 108" and that "[t]ypically <u>a modem connection</u> (not shown) may be used to connect the PC 102 to the packet 108 <u>in a conventional manner</u>". (emphasis added) (See Levin, Column 3, lines 5-10). Additionally, Levin states that "[t]he PC 102 dials into an access server 106 that is connected to the Internet or other database service via a logical network interface (not shown)" and that "[t]he logical network interface may be a local area network (LAN), a Serial Line Internet Protocol (SLIP) connection over a modem, an ISDN port or via a connection to a special LAN such as an ATM LAN or a LAN that offers bandwidth reservation". (See Levin, Column 4, lines 23-29) It is respectfully submitted that none of Levin's statements provides any specific teaching as to mobile appliances or wireless communication. In fact, terms such as "modem connection" and "ISDN port" are typically associated with hard-wired appliances. Thus, Levin does not teach or disclose a <u>method that receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where, in turn, the selected electronic data source from the network server is transmitted to the mobile information appliance of the user</u>. Namely, the scope of Applicants' claims is specifically directed to speech-based navigation via mobile information appliances. This novel concept is not disclosed by the Levin reference and Applicants' claims would not read on the Levin reference.

Therefore, the Applicants respectfully submit that independent claims 56, 65 and 74 are not anticipated by the Levin reference. As such, claims 56, 65 and 74 fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

Claims 57-64, 66-73 and 75-82 depend, either directly or indirectly, from claims 56, 65 and 74 and recite additional features therefor. Since Levin fails to anticipate Applicants' invention as recited in Applicants' independent claims 56, 65 and 74, dependent claims 57-64, 66-73 and 75-82 are also not anticipated under 35 U.S.C. § 102 and are allowable for the same reason noted above.

## <u>Conclusion</u>

Thus, the Applicants submit that all of these claims now fully satisfy the requirements of 35 U.S.C. §102. Consequently, the Applicants believe that all these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

9/19/01

Kin-Wah Tong, Attorney
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
First Floor,
Shrewsbury, New Jersey 07702

GAU 2155

PTO/SB/21 (08-00)

Please type a plus sign (+) inside this box ⟶ ☐+

Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

OIPE JCS
SEP 2 1 2001
PATENT & TRADEMARK OFFICE

# TRANSMITTAL FORM

*(to be used for all correspondence after initial filing)*

| | |
|---|---|
| **Application Number** | 09/608,872 |
| **Filing Date** | June 30, 2000 |
| **First Named Inventor** | HALVERSON |
| Group Art Unit | 2155 |
| Examiner Name | F. BACKER |
| Attorney Docket Number | SRI 1 P 037B |
| Total Number of Pages in This Submission | |

ENCLOSURES *(check all that apply)*

- ☒ Fee Transmittal Form
  - ☒ Fee Attached
- ☒ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☒ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53

- ☐ Assignment Papers *(for an Application)*
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☒ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s)

- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) *(please identify below)*:

RECEIVED
SEP 26 2001
Technology Center 2100

Remarks

## SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

| | |
|---|---|
| Firm *or* Individual name | KIN-WAH TONG |
| Signature | |
| Date | September 19, 2001 |

## CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231 on this date: September 19, 2001

| | | | |
|---|---|---|---|
| Typed or printed name | Linda DeNardi | | |
| Signature | Linda DeNardi | Date | September 19, 2001 |

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be send to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

U.S. Patent and Trademark Office. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no is are required to respond to a collection of information unless it 's a valid OMB control number.

# FEE TRANSMITTAL
# for FY 2001

*Patent fees are subject to annual revision.*

DURING PENDENCY, PLEASE CHARGE DEPOSIT ACCOUNT 20-0782 FOR ANY 37 C.F.R. 1.16 AND/OR 37 C.F.R. 1.17 FEES DUE AND NOT OTHERWISE AUTHORIZED. PLEASE CREDIT DEPOSIT ACCOUNT 20-0782 FOR ANY OVERPAYMENTS

**TOTAL AMOUNT OF PAYMENT** ($) 195.00

| *Complete if Known* | |
|---|---|
| Application Number | 09/608,872 |
| Filing Date | June 30, 2000 |
| First Named Inventor | HALVERSON |
| Examiner Name | F. BACKER |
| Group / Art Unit | 2155 |
| Attorney Docket No. | SRI 1P037B |

OIPE JCGS
SEP 21 2001
PATENT & TRADEMARK OFFICE

RECEIVED
SEP 26 2001
Technology Center 2100

## METHOD OF PAYMENT (check one)

1. ☒ The Commissioner is hereby authorized to charge indicated fees and credit any over payments to:

Deposit Account Number: 20-0782

Deposit Account Name:

☒ Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17

☒ Applicant claims small entity status. See 37 CFR 1.27

2. ☒ Payment Enclosed:

☒ Check ☐ Credit card ☐ Money Order ☐ Other

## FEE CALCULATION

### 1. BASIC FILING FEE

| Large Entity Fee Code | Large Entity Fee ($) | Small Entity Fee Code | Small Entity Fee ($) | Fee Description | Fee Paid |
|---|---|---|---|---|---|
| 101 | 710 | 201 | 355 | Utility filing fee | |
| 106 | 320 | 206 | 160 | Design filing fee | |
| 107 | 490 | 207 | 245 | Plant filing fee | |
| 108 | 710 | 208 | 355 | Reissue filing fee | |
| 114 | 150 | 214 | 75 | Provisional filling fee | |

SUBTOTAL (1) ($) 0

### 2. EXTRA CLAIM FEES

| | | | | Extra Claims | | Fee from below | | Fee Paid |
|---|---|---|---|---|---|---|---|---|
| Total Claims | | -20** | = | 0 | X | | = | 0 |
| Independent Claims | | -3** | = | 0 | X | | = | 0 |
| Multiple Dependent | | | | | X | | = | 0 |

| Large Entity Fee Code | Large Entity Fee ($) | Small Entity Fee Code | Small Entity Fee ($) | Fee Description |
|---|---|---|---|---|
| 103 | 18 | 203 | 9 | Claims in excess of 20 |
| 102 | 80 | 202 | 40 | Independent claims in excess of 3 |
| 104 | 270 | 204 | 135 | Multiple dependent claim, if not paid |
| 109 | 80 | 209 | 40 | ** Reissue independent claims over original patent |
| 110 | 18 | 210 | 9 | ** Reissue claims in excess of 20 and over original patent |

SUBTOTAL (2) ($) 0

**or number previously paid, if greater; For Reissues, see above

## FEE CALCULATION (continued)

### 3. ADDITIONAL FEES

| Large Entity Fee Code | Large Entity Fee ($) | Small Entity Fee Code | Small Entity Fee ($) | Fee Description | Fee Paid |
|---|---|---|---|---|---|
| 105 | 130 | 205 | 65 | Surcharge - late filing fee or oath | |
| 127 | 50 | 227 | 25 | Surcharge - late provisional filing fee or cover sheet. | |
| 139 | 130 | 139 | 130 | Non-English specification | |
| 147 | 2,520 | 147 | 2,520 | For filing a request for reexamination | |
| 112 | 920* | 112 | 920* | Requesting publication of SIR prior to Examiner action | |
| 113 | 1,840* | 113 | 1,840* | Requesting publication of SIR after Examiner action | |
| 115 | 110 | 215 | 55 | Extension for reply within first month | |
| 116 | 390 | 216 | 195 | Extension for reply within second month | 195.00 |
| 117 | 890 | 217 | 445 | Extension for reply within third month | |
| 118 | 1,390 | 218 | 695 | Extension for reply within fourth month | |
| 128 | 1,890 | 228 | 945 | Extension for reply within fifth month | |
| 119 | 310 | 219 | 155 | Notice of Appeal | |
| 120 | 310 | 220 | 155 | Filing a brief in support of an appeal | |
| 121 | 270 | 221 | 135 | Request for oral hearing | |
| 138 | 1,510 | 138 | 1,510 | Petition to institute a public use proceeding | |
| 140 | 110 | 240 | 55 | Petition to revive – unavoidable | |
| 141 | 1,240 | 241 | 620 | Petition to revive – unintentional | |
| 142 | 1,240 | 242 | 620 | Utility issue fee (or reissue) | |
| 143 | 440 | 243 | 220 | Design issue fee | |
| 144 | 600 | 244 | 300 | Plant issue fee | |
| 122 | 130 | 122 | 130 | Petitions to the Commissioner | |
| 123 | 130 | 123 | 130 | Petitions related to provisional applications | |
| 126 | 180 | 126 | 180 | Submission of Information Disclosure Stmt | |
| 581 | 40 | 581 | 40 | Recording each patent assignment per property (times number of properties) | |
| 146 | 710 | 246 | 355 | Filing a submission after final rejection (37 CFR § 1.129(a)) | |
| 149 | 710 | 249 | 355 | For each additional invention to be examined (37 CFR § 1.129(b)) | |
| 179 | 710 | 279 | 355 | Request for Continued Examination (RCE) | |
| 169 | 900 | 169 | 900 | Request for expedited examination of a design application | |

Other fee (specify)

*Reduced by Basic Filing Fee Paid SUBTOTAL (3) ($) 195.00

| SUBMITTED BY | | | | Complete *(if applicable)* | |
|---|---|---|---|---|---|
| *Name (Print/Type)* | KIN-WAH TONG | *Registration No. Attorney/Agent)* | 39,400 | *Telephone* | (732) 530-9404 |
| *Signature* | | | | *Date* | SEPTEMBER 19, 2001 |

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

#13

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

| APPLICATION NUMBER | FILING DATE | FIRST NAMED APPLICANT | ATTY. DOCKET NO./TITLE |
|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B |

**CONFIRMATION NO. 2382**

*OC000000006829467*

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

Date Mailed: 10/02/2001

## NOTICE REGARDING POWER OF ATTORNEY

This is in response to the Power of Attorney filed 09/21/2001.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

---

Lavinia Johnson

LAVINIA D JOHNSON
2100 7033085229

OFFICE COPY

#11

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

| APPLICATION NUMBER | FILING DATE | FIRST NAMED APPLICANT | ATTY. DOCKET NO./TITLE |
|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B |

**CONFIRMATION NO. 2382**

*OC000000006829442*

*OC000000006829442*

C. DOUGLAS McDONALD, ESQ.
CALTON FIELDS, et al.
P. O. BOX 3239
TAMPA,, FL 33601-3239

Date Mailed: 10/02/2001

## NOTICE REGARDING POWER OF ATTORNEY

This is in response to the Power of Attorney filed 09/21/2001.

- The Power of Attorney to you in this application has been revoked by the assignee who has intervened as provided by 37 CFR 3.71. Future correspondence will be mailed to the new address of record(37 CFR 1.33).

---

Lavinia Johnson
LAVINIA D JOHNSON
2100 7033085229

OFFICE COPY

**UNITED STA[illegible] DEPARTMENT OF COMMERCE**
**Patent and Trademark Office**
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

25

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. |
|---|---|---|---|
| 09/608,872 | 06/30/00 | HALVERSEN C | SRILP037B |

TM02/1010

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY NJ 07702

| EXAMINER | |
|---|---|
| PACKER, F | |
| **ART UNIT** | **PAPER NUMBER** |
| 2155 | 14 |

**DATE MAILED:** 10/10/01

**Please find below and/or attached an Office communication concerning this application or proceeding.**

**Commissioner of Patents and Trademarks**

PTO-90C (Rev. 2/95)
*U.S. GPO: 2000-473-000/44602

1- File Copy 25

| | | | |
|---|---|---|---|
| ***Office Action Summary*** | **Application No.**<br>09/608,872 | **Applicant(s)**<br>HALVERSEN ET AL. | |
| | **Examiner**<br>Firmin Backer | **Art Unit**<br>2155 | |

***-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --***

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136 (a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on 26 September 2001 .

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) 56-82 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 56-82 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claims _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are objected to by the Examiner.

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved.

12)☐ The oath or declaration is objected to by the Examiner.

**Priority under 35 U.S.C. § 119**

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

**Attachment(s)**

15) ☐ Notice of References Cited (PTO-892)

16) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

17) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .

18) ☐ Interview Summary (PTO-413) Paper No(s). _____ .

19) ☐ Notice of Informal Patent Application (PTO-152)

20) ☐ Other: .

U.S. Patent and Trademark Office
PTO-326 (Rev. 01-01)

**Office Action Summary**

Part of Paper No. 4

*Response to Request for Reconsideration*

This is in response to a request for reconsideration file on September 26th, 2001. Claims 56-82 are being reconsidered in this action.

## *Double Patenting*

1. A rejection based on double patenting of the "same invention" type finds its support in the language of 35 U.S.C. 101 which states that "whoever invents or discovers any new and useful process ... may obtain a patent therefor ..." (Emphasis added). Thus, the term "same invention," in this context, means an invention drawn to identical subject matter. See *Miller v. Eagle Mfg. Co.*, 151 U.S. 186 (1894); *In re Ockert*, 245 F.2d 467, 114 USPQ 330 (CCPA 1957); and *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970).

A statutory type (35 U.S.C. 101) double patenting rejection can be overcome by canceling or amending the conflicting claims so they are no longer coextensive in scope. The filing of a terminal disclaimer cannot overcome a double patenting rejection based upon 35 U.S.C. 101.

2. Claims 56-82 are provisionally rejected under 35 U.S.C. 101 as claiming the same invention as that of claims 56-126 of copending Application No. 09/524,095. Although the conflicting claims are not identical, they are not patentably distinct. It would have been obvious to one of ordinary skill in the art to observed that the omission of the limitations "**soliciting additional input from the user, including user interaction in a modality different that the original request and, refining the navigation query, based upon the additional input**", of applicant claims 56-82 are already in the Co-pending application 09/524,095, as such they are obvious variation of the inventive concept defined in claims 56-126 of the Co-pending application 09/524,095. See In re Karlson, 136USPQ 184 (CCPA 1963). This is a provisional double patenting rejection since the conflicting claims have not in fact been patented.

## *Claim Rejections - 35 USC § 102*

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent.

4. Claims 56-82 are rejected under 35 U.S.C. 102(e) as being anticipated by Levin et al. (U.S. Patent No. 6,173,279).

5. As per claim 56, Levin et al teach a method for speech-based navigation (*information server, 110*) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising receiving a spoken request (*receive a natural language query*) for desired information from the user (*user, 112*) utilizing the mobile information appliance (*PC, 102*) of the user; rendering an interpretation (*creating a semantic representation*) of the spoken request, constructing a navigation (*generating search)* query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting (*sending*) the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)

6. As per claim 57, 58, 62-64, Levin et al teach a method of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

7. As per claim 59, Levin et al teach a method of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

8. As per claim 60, Levin et al teach a method wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

9. As per claim 61, Levin et al teach a method wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

10. As per claim 65, Levin et al teach a computer system for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising a code segment receiving a

spoken request (*receive a natural language query*) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user; a code segment rendering an interpretation (*creating a semantic representation*) of the spoken request, a code segment constructing a navigation (*generating search)* query based upon the interpretation; a code segment utilizing the navigation query to select a portion of the electronic data source; and a code segment transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)

11. As per claim 66, 67, 71-73, Levin et al teach a system of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

12. As per claim 68, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

13. As per claim 69, Levin et al teach a system wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

14. As per claim 70, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

15. As per claim 74, Levin et al teach a system for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising receiving a spoken request (*receive a natural language query*) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user; rendering an interpretation (*creating a semantic representation*) of the spoken request, constructing a navigation (*generating search)* query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)

16. As per claim 75, 76, 80-81, Levin et al teach a method of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

17. As per claim 77, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the

navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

18. As per claim 78, Levin et al teach a system wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

19. As per claim 79, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

### *Response to Arguments*

1. Applicant's arguments filed on September 26th, 2001 have been fully considered but they are not persuasive. ***

a. Applicant argues that the statutory-type obviousness double patenting is not appropriate. Examiner respectfully disagrees with applicant characterization of the statutory-type obviousness double patenting concept. The inventive concepts in the applications are not patenbly different. Different variation of the same inventive concept is being claimed twice. According to MPEP in determining whether a statutory basis for a double patenting rejection exists, the question to be asked is: Is the same invention being claimed twice? 35 U.S.C. 101 prevents two patents from issuing on the same invention. "Same invention" means identical subject matter. Miller v. Eagle Mfg. Co., 151 U.S.

186 (1984); In re Vogel, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and In re Ockert, 245 F.2d 467, 114 USPQ 330 (CCPA 1957).

b. Applicant further argues that the prior art "fails to teach or suggest the novel concept of speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where in turn the selected electronic data source from the network server is transmitted to the mobile information appliance of the user." Examiner respectfully disagrees with the applicant perspective and characterization of Levin inventive concept. Levin teach that use of a personal computer, a user establishes connection with a network. In the field of the network communication, a personal computer is not limited to desktop, but also handheld computer as well as laptop which are considered to be mobile appliances. In Levin inventive concept, an information server 110 receives natural language which is the same as spoken word. One the natural language query is process, the service host then transmit the result of the query to the pc. (see column 3 lines 5-35, 6 lines 25-59).

## *Conclusion*

**THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Firmin Backer whose telephone number is 703-305-0624. The examiner can normally be reached on Mon-Thu 8:30-6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Sheikh Ayaz can be reached on 703-305-9648. The fax phone numbers for the organization where this application or proceeding is assigned are 703-305-3718 for regular communications and 703-305-5352 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

Firmin Backer
October 2, 2001

AYAZ SHEIKH
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

UNITED STATES F NT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B | 2382 |

7590 01/16/2002

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER | |
|---|---|
| BACKER, FIRMIN | |
| **ART UNIT** | **PAPER NUMBER** |
| 2155 | 15 |

DATE MAILED: 01/16/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 07-01)

## *Interview Summary*

| | | |
|---|---|---|
| **Application No.** | **Applicant(s)** | |
| 09/608,872 | HALVERSEN ET AL. | |
| **Examiner** | **Art Unit** | |
| Firmin Backer | 2155 | |

All participants (applicant, applicant's representative, PTO personnel):

(1) *Firmin Backer*.

(2) *Ario Etienne*.

(3) *Kin-Wah Tong*.

(4) _____.

Date of Interview: *08 January 2002* .

Type: a)☒ Telephonic b)☐ Video Conference
c)☐ Personal [copy given to: 1)☐ applicant 2)☐ applicant's representative]

Exhibit shown or demonstration conducted: d)☐ Yes e)☐ No.
If Yes, brief description: _____ .

Claim(s) discussed: *56* .

Identification of prior art discussed: *6,173,279* .

Agreement with respect to the claims f)☐ was reached. g)☐ was not reached. h)☐ N/A.

Substance of Interview including description of the general nature of what was agreed to if an agreement was reached, or any other comments: *Applicant argues that the statutory double patenting rejection is improper and should be withdrawn. Applicant argues that the prior art fails to teach all the limitations of the inventive concept especially the use of wireless communication...* .

(A fuller description, if necessary, and a copy of the amendments which the examiner agreed would render the claims allowable, if available, must be attached. Also, where no copy of the amendments that would render the claims allowable is available, a summary thereof must be attached.)

i)☐ It is not necessary for applicant to provide a separate record of the substance of the interview(if box is checked).

Unless the paragraph above has been checked, THE FORMAL WRITTEN REPLY TO THE LAST OFFICE ACTION MUST INCLUDE THE SUBSTANCE OF THE INTERVIEW. (See MPEP Section 713.04). If a reply to the last Office action has already been filed, APPLICANT IS GIVEN ONE MONTH FROM THIS INTERVIEW DATE TO FILE A STATEMENT OF THE SUBSTANCE OF THE INTERVIEW. See Summary of Record of Interview requirements on reverse side or on attached sheet.

Examiner Note: You must sign this form unless it is an Attachment to a signed Office action.

Examiner's signature, if required

#16/C
B, 1/17/02
Enter 2-8-02

09/608,872

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

**PATENT APPLICATION**

RECEIVED
JAN 10 2002
Technology Center 2100

Applicant: **Halverson et al.**

Case: **SRI1P037B**

Serial No.: **09/608,872** Filed: **June 30, 2000**

Group Art Unit: **2155**

Examiner: **Firmin Backer**

Title: **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT**

ASSISTANT COMMISSIONER FOR PATENTS
**Box AF**
Washington, D. C. 20231

S I R:

## RESPONSE UNDER 37 C.F.R. § 1.116

This response addresses the Final Office Action dated October 10, 2001 (Paper No. 14).

IN THE CLAIMS

**Please amend claims 56 and 65 as shown below. These claims are "clean version" of the amended claims, i.e., with changes incorporated into the claims, whereas the Appendix to this Amendment illustrates the amended claims using underlines and brackets to indicate addition and deletion, respectively.**

56. (Amended) A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:

Sub B1 C1

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d) utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication.

Sub B2 C2

65. (Amended) A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication.

## REMARKS

Applicants' representative would like to thank Examiner Backer and Primary Examiner Etienne for kindly taking a substantial amount of time on January 8, 2002 to discuss the merits of the subject invention. Applicants' representative is aware of the time constraint that is placed on the Examiners and is appreciative of the Examiners' willingness to devote such large quantity of time to discuss the case on the merit.

In view of the following discussion, the Applicants submit that none of the claims now pending in the application are anticipated under the provisions of 35 U.S.C. § 102. Thus, the Applicants believe that all of these claims are now in allowable form.

## I. REJECTION OF CLAIMS 56-82 UNDER DOUBLE PATENTING

The Examiner provisionally rejected claims 56-82 in Paragraphs 1-2 of the Final Office Action based on statutory type double patenting under 35 U.S.C. § 101 as claiming the same invention as that of claims 56-126 of copending Application No. 09/524,095. Applicants respectfully traverse the rejection.

First, the Examiner noted that "it would have been obvious to one of ordinary skill in the art to observe that the omission of the limitations '**soliciting additional input from the user, including user interaction in a modality different tha[n] the original request and, refining the navigation query, based upon the additional input**'. After noting the differences between the scope of the claims between the two applications, the Examiner then concluded that claims 56-82 "are obvious variation of the inventive concept defined in claims 56-126 of co-pending application 09/524,095".

Pursuant to the Examiner Interview, Applicants again directed Examiner's attention to the fact that there are two types of double patenting rejections: "statutory" and "non-statutory (obviousness-type)". MPEP 804 states that "[i]n determining whether a statutory basis for a double patenting rejection exists, the question to be asked is: Is the same invention being claimed twice?" "A reliable test for double patenting under 35 U.S.C. 101 is whether a claim in the application could be literally infringed without literally infringing a corresponding claim in the patent". Given the substantial differences between the claims of the two applications as noted by the Examiner, Applicants respectfully submit that applying the statutory double patenting test as promoted in the MPEP would not produce a statutory double patenting rejection in the present application.

Second, it should be noted that the present application is a continuation of the co-pending application 09/524,095. As such, if and when these two applications mature into issued patents, both patents will have the same term.

As such, Applicants submit that the present statutory double patenting rejection against claims 56-82 is inappropriate. The Examiners indicated that they will reconsider the present statutory type double patenting under 35 U.S.C. § 101.

## II. REJECTION OF CLAIMS 56-82 UNDER 35 U.S.C. § 102

The Examiner has rejected claims 56-82 in Paragraphs 4-19 of the Final Office Action as being anticipated by the Levin et al. patent (US Patent 6,173,279 issued January 9, 2001, hereinafter referred to as Levin). The rejection is respectfully traversed.

Levin teaches "a method of using at least one natural language query to retrieve information from one or more data resources and further performing a requested action using the retrieved information is disclosed". (See Levin, Column 2, lines 15-18) Namely, Levin teaches a method for using natural language query to obtain information, where upon receipt of the requested information, a desired action is executed based upon the requested information. To illustrate, Levin provides the example, where a user employs natural language to request the telephone number of a restaurant. Upon receipt of the telephone number, the telephone number is actually dialed for the user. (See Levin, Column 3 line 62 to Column 4, line 1)

In contrast, Levin fails to teach or suggest the novel concept of speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where, in turn, the selected electronic data source from the network server is transmitted to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication. Specifically, Applicants' independent claims 56, 65 and 74 positively recite:

56. A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d)utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication. (emphasis added)

65. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication. (emphasis added)

74. A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

(a) a mobile information appliance operable to receive a spoken request for desired information from the user;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

> (e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of a data link of the electronic communications infrastructure between a mobile information appliance of the user and the one or more network servers utilizes wireless communication. (emphasis added)

Applicants' invention teaches a novel method and apparatus for speech-based navigation where the method receives spoken request for desired information from the user utilizing the **mobile information appliance** of the user and where, in turn, the selected electronic data source from the network server is **transmitted to the mobile information appliance** of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication. Specifically, Applicants address the criticality of providing speech-based navigation via a mobile, i.e., wireless communication, approach in addition to spoken natural language. It has been noted that with the proliferation of various mobile appliances, it would be advantageous to allow these mobile appliances to access the same vastness of electronic data sources that are available to hard-wired appliances like a desktop computer. However, the very essence of a mobile appliance is its portability, small size and ease of use. As such, unlike hard-wired appliances, mobile appliances are not equipped with large bulky input devices. In fact, even if the mobile appliance is equipped with extensive input devices, most users would still find these "shrunken" input devices to be cumbersome and difficult to use, e.g., an electronic representation of a keyboard on a PDA and the like.

To further exacerbate the problem, obtaining information from an electronic data source may require extensive and complex interaction between the user's mobile appliance and the system holding the electronic data source. Thus, the limited or cumbersome input/output capability of a mobile appliance presents a substantial barrier to its ability to access a data resource that requires extensive and complex interaction.

In contrast, Levin teaches that "[u]sing a personal computer (PC) 102, a user establishes a connection with packet network 108 via an access server 106". Levin then states that "[t]he user may also use a telephone 103 to connect to the packet network 108" and that "[t]ypically a modem connection (not shown) may be used to connect the PC 102 to the packet 108 in a conventional manner". (emphasis added) (See Levin, Column 3, lines 5-10). Additionally, Levin states that "[t]he PC 102 dials into an access server 106 that is connected to the Internet or other database service via a logical network interface (not shown)" and that "[t]he logical network interface may be a local area network (LAN), a Serial Line Internet Protocol (SLIP) connection over a modem, an ISDN port or via a connection to a special LAN such as an ATM LAN or a LAN that offers bandwidth reservation". (See Levin, Column 4, lines 23-29) It is respectfully submitted that none of Levin's statements provides any specific teaching as to mobile appliances or wireless communication. In fact, terms such as "modem connection" and "ISDN port" are typically associated with hard-wired appliances. Thus, Levin does not teach or disclose a method that receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where, in turn, the selected electronic data source from the network server is transmitted to the mobile information appliance of the user via wireless communication over at least a portion of the data link. Namely, the scope of Applicants' claims is specifically directed to speech-based navigation via mobile information appliances. This novel concept is not disclosed by the Levin reference and Applicants' claims would not read on the Levin reference.

Pursuant to the Examiner Interview, Applicants have agreed to incorporate the term " wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication", into the body of the independent claims. This term previously existed in the preamble of the independent claims. Thus, since this term previously existed in the originally filed independent claims, the present amendment is not implemented in view of the cited prior art. In fact, Applicants take the position that the scope of the independent claims did not change as a result of this amendment and that this amendment served to clarify the claims to the Examiner's satisfaction.

Additionally, it should be noted that no amendment was applied to independent claim 74, since the above-identified term is already in the body of the independent claim

74.

Therefore, the Applicants respectfully submit that independent claims 56, 65 and 74 are not anticipated by the Levin reference. As such, claims 56, 65 and 74 fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

Claims 57-64, 66-73 and 75-82 depend, either directly or indirectly, from claims 56, 65 and 74 and recite additional features therefor. Since Levin fails to anticipate Applicants' invention as recited in Applicants' independent claims 56, 65 and 74, dependent claims 57-64, 66-73 and 75-82 are also not anticipated under 35 U.S.C. § 102 and are allowable for the same reason noted above.

## Conclusion

Thus, the Applicants submit that all of these claims now fully satisfy the requirements of 35 U.S.C. §102. Consequently, the Applicants believe that all these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the maintenance of the present final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

1/10/02

Kin-Wah Tong, Attorney
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
First Floor,
Shrewsbury, New Jersey 07702

## Appendix
## (Marked-up copy of amended claims)

56. (Amended) A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein [at least a portion of] a data link is established between a mobile information appliance of the user and the one or more network servers [utilize wireless communication], comprising the steps of:

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d)utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication.

65. (Amended) A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein [at least a portion of] a data link is established between a mobile information appliance of the user and the one or more network servers [utilizes wireless communication], comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user;

(b) a code segment that renders an interpretation of the spoken request.

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication.

# TELEFAX COVER SHEET

Official

**MOSER, PATTERSON & SHERIDAN**
ATTORNEYS AT LAW
595 SHREWSBURY AVENUE
FIRST FLOOR
SHREWSBURY, NJ 07702
**TELEPHONE (732) 530-9404**
**TELEFAX (732) 530-9808**

1/10/02

**************************************************

THIS TELEFAX MESSAGE IS ADDRESSED TO THE PERSON OR COMPANY LISTED BELOW. IF IT WAS SENT OR RECEIVED INCORRECTLY, OR YOU ARE NOT THE INTENDED RECIPIENT, PLEASE TAKE NOTICE THAT THIS MESSAGE MAY CONTAIN PRIVILEGED OR CONFIDENTIAL MATERIAL, AND YOUR DUE REGARD FOR THIS INFORMATION IS NECESSARY. YOU MAY ARRANGE TO RETURN THIS MATERIAL BY CALLING THE FIRM LISTED ABOVE AT (732) 530-9404

**************************************************

THIS MESSAGE HAS 13 PAGES INCLUDING THIS SHEET

TO: Assistant Commissioner of Patents

FAX NO.: 703-746-7238

FROM: Kin-Wah Tong

DATE: January 10, 2002

MATTER: Serial No. 09/608,872 Filed: June 30, 2000

DOCKET NO.: SRI 1P037B

APPLICANT: HALVERSON, et al

The following has been received in the U.S. Patent and Trademark Office on the date of this facsimile:

___ Petition
___ Disclosure Statement & PTO-1449
___ Priority Document
___ Drawings (____ sheets) informal
X Response Under 37 CFR 1.116
___ ________________
___ ________________

X Transmittal Letter (2 copies)
___ Fee Transmittal (2 copies)
___ Deposit Account Transaction
X Facsimile Transmission Certificate dated January 10, 2002
___ ________________
___ ________________

## CERTIFICATE OF TRANSMISSION UNDER 37 C.F.R. §1.6

I hereby certify that this correspondence is being transmitted by facsimile to the Assistant Commissioner for Patents, Box AF, Washington, DC 20231 on January 10, 2002.
Facsimile No. 703-746-7238.

Linda DeNardi
Name of person signing this certificate

Linda DeNardi January 10, 2002
Signature and date

Received from < 732 530 9808 > at 1/10/02 4:06:40 PM [Eastern Standard Time]

PTO/SB/21 (08-00)
Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Please type a plus sign (+) inside this box → ☐ +

# TRANSMITTAL FORM

*(to be used for all correspondence after initial filing)*

| | |
|---|---|
| Application Number | 09/608,872 |
| Filing Date | June 30, 2000 |
| First Named Inventor | HALVERSON |
| Group Art Unit | 2155 |
| Examiner Name | F. BACKER |
| Attorney Docket Number | SRI 1 P 037B |

Total Number of Pages in This Submission: 13

## ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☒ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Assignment Papers *(for an Application)*
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s)
- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) *(please identify below)*:

Remarks: **It is believed no fee is due. However, in the event a fee is due, kindly charge that fee to deposit account number 20-0782. To facilitate that charge, a duplicate copy of this letter is enclosed**

## SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

| | |
|---|---|
| Firm or Individual name | PATRICIA A. VERLANGIERI, Reg. No. 42,201 |
| Signature | Patricia A Verlangieri |
| Date | January 10, 2001 |

PTO/SB/21 (08-00)
Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Please type a plus sign (+) inside this box → ☐+

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

# TRANSMITTAL FORM

*(to be used for all correspondence after initial filing)*

| | |
|---|---|
| **Application Number** | 09/608,872 |
| **Filing Date** | June 30, 2000 |
| **First Named Inventor** | HALVERSON |
| Group Art Unit | 2155 |
| Examiner Name | F. BACKER |
| Attorney Docket Number | SRI 1 P 037B |

Total Number of Pages in This Submission: 13

## ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☒ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Assignment Papers *(for an Application)*
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s)
- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) *(please identify below)*:

Remarks: **It is believed no fee is due. However, in the event a fee is due, kindly charge that fee to deposit account number 20-0782. To facilitate that charge, a duplicate copy of this letter is enclosed**

## SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

| | |
|---|---|
| Firm or Individual name | PATRICIA A. VERLANGIERI, Reg. No. 42,201 |
| Signature | Patricia A Verlangieri |
| Date | January 10, 2001 |

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be send to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

UNITED STAT… PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B | 2382 |

7590 01/28/2002

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER |
|---|
| BACKER, FIRMIN |

| ART UNIT | PAPER NUMBER |
|---|---|
| 2155 | 17 |

DATE MAILED: 01/28/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 07-01)

| **Advisory Action** | **Application No.**<br>09/608,872 | **Applicant(s)**<br>HALVERSEN ET AL. | |
|---|---|---|---|
| | **Examiner**<br>Firmin Backer | **Art Unit**<br>2155 | |

***--The MAILING DATE of this communication appears on the cover sheet with the correspondence address --***

THE REPLY FILED 17 January 2002 FAILS TO PLACE THIS APPLICATION IN CONDITION FOR ALLOWANCE. Therefore, further action by the applicant is required to avoid abandonment of this application. A proper reply to a final rejection under 37 CFR 1.113 may only be either: (1) a timely filed amendment which places the application in condition for allowance; (2) a timely filed Notice of Appeal (with appeal fee); or (3) a timely filed Request for Continued Examination (RCE) in compliance with 37 CFR 1.114.

PERIOD FOR REPLY [check either a) or b)]

a) ☒ The period for reply expires 3 months from the mailing date of the final rejection.

b) ☐ The period for reply expires on: (1) the mailing date of this Advisory Action, or (2) the date set forth in the final rejection, whichever is later. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of the final rejection. ONLY CHECK THIS BOX WHEN THE FIRST REPLY WAS FILED WITHIN TWO MONTHS OF THE FINAL REJECTION. See MPEP 706.07(f).

Extensions of time may be obtained under 37 CFR 1.136(a). The date on which the petition under 37 CFR 1.136(a) and the appropriate extension fee have been filed is the date for purposes of determining the period of extension and the corresponding amount of the fee. The appropriate extension fee under 37 CFR 1.17(a) is calculated from: (1) the expiration date of the shortened statutory period for reply originally set in the final Office action; or (2) as set forth in (b) above, if checked. Any reply received by the Office later than three months after the mailing date of the final rejection, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

1.☐ A Notice of Appeal was filed on _____. Appellant's Brief must be filed within the period set forth in 37 CFR 1.192(a), or any extension thereof (37 CFR 1.191(d)), to avoid dismissal of the appeal.

2.☒ The proposed amendment(s) will not be entered because:

(a) ☒ they raise new issues that would require further consideration and/or search (see NOTE below);

(b) ☐ they raise the issue of new matter (see Note below);

(c) ☐ they are not deemed to place the application in better form for appeal by materially reducing or simplifying the issues for appeal; and/or

(d) ☐ they present additional claims without canceling a corresponding number of finally rejected claims.

NOTE: *See Continuation Sheet*.

3.☐ Applicant's reply has overcome the following rejection(s): _____.

4.☐ Newly proposed or amended claim(s) _____ would be allowable if submitted in a separate, timely filed amendment canceling the non-allowable claim(s).

5.☐ The a)☐ affidavit, b)☐ exhibit, or c)☐ request for reconsideration has been considered but does NOT place the application in condition for allowance because: _____.

6.☐ The affidavit or exhibit will NOT be considered because it is not directed SOLELY to issues which were newly raised by the Examiner in the final rejection.

7.☒ For purposes of Appeal, the proposed amendment(s) a)☐ will not be entered or b)☐ will be entered and an explanation of how the new or amended claims would be rejected is provided below or appended.

The status of the claim(s) is (or will be) as follows:

Claim(s) allowed: _____.

Claim(s) objected to: _____.

Claim(s) rejected: *56-82*.

Claim(s) withdrawn from consideration: _____.

8.☐ The proposed drawing correction filed on _____ is a)☐ approved or b)☐ disapproved by the Examiner.

9.☐ Note the attached Information Disclosure Statement(s)( PTO-1449) Paper No(s). _____.

10.☐ Other: _____

U.S. Patent and Trademark Office
PTO-303 (Rev. 04-01) **Advisory Action** Part of Paper No. 4

Continuation of 2. NOTE: The proposed amendments will not be entered because the raised new issue such as in claims 56 and 65 "wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network utilizes wireless communication" that require further search and/or consideration .

AYAZ SHEIKH
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

Official
2/8/02

PTO/SB/30 (08-00)
Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

# REQUEST FOR CONTINUED EXAMINATION (RCE) TRANSMITTAL

**Subsection (b) of 35 U.S.C. § 132, effective on May 29, 2000, provides for continued examination of an utility or plant application filed on or after June 8, 1995. See The American Inventors Protection Act of 1999 (AIPA).**

| | |
|---|---|
| *Application Number* | 09/608,872 |
| *Filing Date* | June 30, 2000 |
| *Examiner Name* | F. Backer |
| *First Named Inventor* | Halversen |
| *Group Art Unit* | 2155 |
| *Attorney Docket Number* | SRI 1P037B |

#18
2-8-02

This is a Request for Continued Examination (RCE) under 37 C.F.R. § 1.114 of the above-identified application.
***NOTE:*** *37 C.F.R. § 1.114 is effective on May 29, 2000. If the above-identified application was filed prior to May 29, 2000, applicant may wish to consider filing a continued prosecution application (CPA) under 37 C.F.R. § 1.53 (d) (PTO/SB/29) instead of a RCE to be eligible for* the patent term adjustment provisions of the AIPA. See Changes to Application Examination and Provisional Application Practice, Interim Rule, 65 *Fed. Reg.* 14865 (Mar. 20, 2000), 1233 *Off. Gaz. Pat. Office* 47 (Apr. 11, 2000), which established RCE practice.

1. Submission required under 37 C.F.R. § 1.114
   - a. ☒ Previously submitted
     - i. ☒ Consider the amendment(s)/reply under 37 C.F.R. § 1.116 previously filed on 1/10/02 (Any unentered amendment(s) referred to above will be entered).
     - ii. ☐ Consider the arguments in the Appeal Brief or Reply Brief previously filed on _____
     - iii. ☐ Other _____
   - b. Enclosed
     - i. ☐ Amendment/Reply
     - ii ☐ Affidavit(s)/Declaration(s)
     - iii. ☐ Information Disclosure Statement (IDS)
     - iv. ☐ Other _____
2. Miscellaneous
   - a. ☐ Suspension of action on the above-identified application is requested under 37 C.F.R. § 1.103(c) for a period of _____months. (Period of suspension shall not exceed 3 months; Fee under 37 C.F.R. § 1.17(i) required)
   - b. ☒ Other Extension Request and Fee Transmittal Sheet
3. Fees The RCE fee under 37 C.F.R. § 1.17(e) is required by 37 C.F.R. § 1.114 when the RCE is filed.
   - a. ☒ The Director is hereby authorized to charge the following fees, or credit any overpayments, to Deposit Account No.20-0782
     - i. ☒ RCE fee required under 37 C.F.R. § 1.17(e)
     - ii. ☒ Extension of time fee (37 C.F.R. §§ 1.136 and 1.17)
     - iii. ☐ Other
   - b. ☐ Check in the amount of $_____ enclosed
   - c. ☐ Payment by credit card (Form PTO-2038 enclosed)

***SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT REQUIRED***

| | | | |
|---|---|---|---|
| *Name (Print/Type)* | KIN-WAH TONG | *Registration No. (Attorney/Agent)* | 39,400 |
| *Signature* | | *Date* | February 8, 2002 |

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND Fees and Completed Forms to the following address: Commissioner for Patents, Box RCE, Washington, DC 20231.

02/12/2002 DHALL1 00000003 200782 09608872

01 FC:215 55.00 CH
02 FC:279 370.00 CH

Received from < 732 530 9808 > at 2/8/02 3:49:42 PM [Eastern Standard Time]

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

# PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)

Docket Number (Optional) SRI 1P037B

In re Application of HALVERSEN

Application Number 09/608,872

Filed June 30, 2000

For Mobile Navigation of Network-Based Electronic Information Using Spoken Input

Group Art Unit 2155

Examiner F. Backer

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a response in the above identified application.

The requested extension and appropriate non-small-entity fee are as follows (check time period desired):

| | |
|---|---|
| ☒ One month (37 CFR 1.17(a)(1)) | $110.00 |
| ☐ Two months (37 CFR 1.17(a)(2)) | $ |
| ☐ Three months (37 CFR 1.17(a)(3)) | $ |
| ☐ Four months (37 CFR 1.17(a)(4)) | $ |
| ☐ Five months (37 CFR 1.17(a)(5)) | $ |

☒ Applicant claims small entity status. See 37 CFR 1.27. Therefore, the fee amount shown above is reduced by one-half, and the resulting fee is: $ 55.00 .

☐ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Commissioner has already been authorized to charge fees in this application to a Deposit Account.

☒ The Commissioner is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 20-0782 .

I have enclosed a duplicate copy of this sheet.

I am the ☐ applicant/inventor.

☐ assignee of record of the entire interest. See 37 CFR 3.71

Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).

☒ attorney or agent of record.

☐ attorney or agent under 37 CFR 1.34(a).

Registration number if acting under 37 CFR 1.34(a). ______ .

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

February 8, 2002

Date

Signature

KIN-WAH TONG

Typed or printed name

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of ______ forms are submitted.

Burden Hour Statement: This form is estimated to take 0.1 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

# PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)

Docket Number (Optional)
SRI 1P037B

In re Application of HALVERSEN

Application Number 09/608,872

Filed June 30, 2000

For Mobile Navigation of Network-Based Electronic Information Using Spoken Input

Group Art Unit 2155

Examiner F. Backer

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a response in the above identified application.

The requested extension and appropriate non-small-entity fee are as follows (check time period desired):

| | |
|---|---|
| ☒ One month (37 CFR 1.17(a)(1)) | $110.00 |
| ☐ Two months (37 CFR 1.17(a)(2)) | $ |
| ☐ Three months (37 CFR 1.17(a)(3)) | $ |
| ☐ Four months (37 CFR 1.17(a)(4)) | $ |
| ☐ Five months (37 CFR 1.17(a)(5)) | $ |

☒ Applicant claims small entity status. See 37 CFR 1.27. Therefore, the fee amount shown above is reduced by one-half, and the resulting fee is: $ 55.00 .

☐ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Commissioner has already been authorized to charge fees in this application to a Deposit Account.

☒ The Commissioner is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 20-0782 .
I have enclosed a duplicate copy of this sheet.

I am the ☐ applicant/inventor.

☐ assignee of record of the entire interest. See 37 CFR 3.71
Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).

☒ attorney or agent of record.

☐ attorney or agent under 37 CFR 1.34(a).
Registration number if acting under 37 CFR 1.34(a). ____ .

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

February 8, 2002
Date

Signature

KIN-WAH TONG
Typed or printed name

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of ____ forms are submitted.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

# FEE TRANSMITTAL for FY 2002

*Patent fees are subject to annual revision.*

**TOTAL AMOUNT OF PAYMENT** ($) 425

| *Complete if Known* | |
|---|---|
| Application Number | 09/608,872 |
| Filing Date | June 30, 2000 |
| First Named Inventor | Halversen |
| Examiner Name | F. Backer |
| Group / Art Unit | 2155 |
| Attorney Docket No. | SRI 1P037B |

**METHOD OF PAYMENT (check one)**

1. ☒ The Commissioner is hereby authorized to charge indicated fees and credit any over payments to:

Deposit Account Number: 20-0782

Deposit Account Name:

☒ Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17

☒ Applicant claims small entity status. See 37 CFR 1.27

2. ☐ Payment Enclosed:

☐ Check ☐ Credit card ☐ Money Order ☐ Other

**FEE CALCULATION**

**1. BASIC FILING FEE**

| Large Entity Fee Code | Large Entity Fee ($) | Small Entity Fee Code | Small Entity Fee ($) | Fee Description | Fee Paid |
|---|---|---|---|---|---|
| 101 | 740 | 201 | 370 | Utility filing fee | |
| 106 | 330 | 206 | 165 | Design filing fee | |
| 107 | 510 | 207 | 255 | Plant filing fee | |
| 108 | 740 | 208 | 370 | Reissue filing fee | |
| 114 | 160 | 214 | 80 | Provisional filing fee | |

SUBTOTAL (1) ($) 0

**2. EXTRA CLAIM FEES**

| | | | Extra Claims | | Fee from below | | Fee Paid |
|---|---|---|---|---|---|---|---|
| Total Claims | | -20 ** = | 0 | X | | = | 0 |
| Independent Claims | | -3 ** = | 0 | X | | = | 0 |
| Multiple Dependent | | | | X | | = | 0 |

| Large Entity Fee Code | Large Entity Fee ($) | Small Entity Fee Code | Small Entity Fee ($) | Fee Description |
|---|---|---|---|---|
| 103 | 18 | 203 | 9 | Claims in excess of 20 |
| 102 | 84 | 202 | 42 | Independent claims in excess of 3 |
| 104 | 280 | 204 | 140 | Multiple dependent claim, if not paid |
| 109 | 84 | 209 | 42 | ** Reissue independent claims over original patent |
| 110 | 18 | 210 | 9 | ** Reissue claims in excess of 20 and over original patent |

SUBTOTAL (2) ($) 0

**or number previously paid, if greater; For Reissues, see above

**FEE CALCULATION (continued)**

**3. ADDITIONAL FEES**

| Large Entity Fee Code | Large Entity Fee ($) | Small Entity Fee Code | Small Entity Fee ($) | Fee Description | Fee Paid |
|---|---|---|---|---|---|
| 105 | 130 | 205 | 65 | Surcharge - late filing fee or oath | |
| 127 | 50 | 227 | 25 | Surcharge - late provisional filing fee or cover sheet. | |
| 139 | 130 | 139 | 130 | Non-English specification | |
| 147 | 2,520 | 147 | 2,520 | For filing a request for reexamination | |
| 112 | 920* | 112 | 920* | Requesting publication of SIR prior to Examiner action | |
| 113 | 1,840* | 113 | 1,840* | Requesting publication of SIR after Examiner action | |
| 115 | 110 | 215 | 55 | Extension for reply within first month | 55.00 |
| 116 | 400 | 216 | 200 | Extension for reply within second month | |
| 117 | 920 | 217 | 460 | Extension for reply within third month | |
| 118 | 1,440 | 218 | 720 | Extension for reply within fourth month | |
| 128 | 1,960 | 228 | 980 | Extension for reply within fifth month | |
| 119 | 320 | 219 | 160 | Notice of Appeal | |
| 120 | 320 | 220 | 160 | Filing a brief in support of an appeal | |
| 121 | 280 | 221 | 140 | Request for oral hearing | |
| 138 | 1,510 | 138 | 1,510 | Petition to institute a public use proceeding | |
| 140 | 110 | 240 | 55 | Petition to revive – unavoidable | |
| 141 | 1,280 | 241 | 640 | Petition to revive – unintentional | |
| 142 | 1,280 | 242 | 640 | Utility issue fee (or reissue) | |
| 143 | 460 | 243 | 230 | Design issue fee | |
| 144 | 620 | 244 | 310 | Plant issue fee | |
| 122 | 130 | 122 | 130 | Petitions to the Commissioner | |
| 123 | 50 | 123 | 50 | Processing fee under 37 CFR 1.17 (q) | |
| 126 | 180 | 126 | 180 | Submission of Information Disclosure Stmt | |
| 581 | 40 | 581 | 40 | Recording each patent assignment per property (times number of properties) | |
| 146 | 740 | 246 | 370 | Filing a submission after final rejection (37 CFR § 1.129(a)) | |
| 149 | 740 | 249 | 370 | For each additional invention to be examined (37 CFR § 1.129(b)) | |
| 179 | 740 | 279 | 370 | Request for Continued Examination (RCE) | 370.00 |
| 169 | 900 | 169 | 900 | Request for expedited examination of a design application | |

Other fee (specify)

*Reduced by Basic Filing Fee Paid SUBTOTAL (3) ($) 425

**SUBMITTED BY** Complete *(if applicable)*

| | | | | | |
|---|---|---|---|---|---|
| *Name (Print/Type)* | KIN-WAH TONG | *Registration No. Attorney/Agent)* | 39,400 | *Telephone* | (732)530-9404 |
| *Signature* | | | | *Date* | FEBRUARY 8, 2002 |

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

# TELEFAX COVER SHEET

**MOSER, PATTERSON & SHERIDAN, LLP**
ATTORNEYS AT LAW
595 SHREWSBURY AVENUE
FIRST FLOOR
SHREWSBURY, NJ 07702
**TELEPHONE (732) 530-9404**
**TELEFAX (732) 530-9808**

Official

RECEIVED

********************************************************

********************************************************

THIS MESSAGE HAS 6 PAGES INCLUDING THIS SHEET

TO: Assistant Commissioner of Patents

FAX NO.: 703-746-7238

FROM: Kin-Wah Tong

DATE: February 8, 2002

MATTER: Serial No. 09/608,872 Filed: June 30, 2000

DOCKET NO.: SRI 1P037B

APPLICANT: HALVERSON, et al

The following has been received in the U.S. Patent and Trademark Office on the date of this facsimile:

- ___ Petition
- ___ Disclosure Statement & PTO-1449
- ___ Priority Document
- ___ Drawings (____ sheets) informal
- X Petition for Extension of Time (2 copies)
- X RCE Transmittal Letter
- X Fee Transmittal (2 copies)
- X Deposit Account Transaction
- X Facsimile Transmission Certificate dated February 8, 2002

## CERTIFICATE OF TRANSMISSION UNDER 37 C.F.R. §1.6

I hereby certify that this correspondence is being transmitted by facsimile to the Assistant Commissioner for Patents, Box AF, Washington, DC 20231 on February 8, 2002, Facsimile No. 703-746-7238.

Linda DeNardi
Name of person signing this certificate

Linda DeNardi February 8, 2002
Signature and date

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B | 2382 |

7590 02/19/2002

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER | |
|---|---|
| BACKER, FIRMIN | |
| ART UNIT | PAPER NUMBER |
| 2155 | 19 |

DATE MAILED: 02/19/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 07-01)

| Office Action Summary | Application No. | Applicant(s) |
|---|---|---|
| | 09/608,872 | HALVERSEN ET AL. |
| | Examiner | Art Unit |
| | Firmin Backer | 2155 |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136 (a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *08 February 2002* .

2a)☐ This action is **FINAL**.  2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *56-82* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *56-82* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claims _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are objected to by the Examiner.

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved.

12)☐ The oath or declaration is objected to by the Examiner.

**Priority under 35 U.S.C. § 119**

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

**Attachment(s)**

15) ☐ Notice of References Cited (PTO-892)

16) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

17) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .

18) ☐ Interview Summary (PTO-413) Paper No(s). _____ .

19) ☐ Notice of Informal Patent Application (PTO-152)

20) ☐ Other:

### *Continued Examination Under 37 CFR 1.114*

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on February 8th, 2002 has been entered.

### *Double Patenting*

2. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. See *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and, *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground provided the conflicting application or patent is shown to be commonly owned with this application. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

3. Claims 56-82 are provisionally rejected under the judicially created doctrine of double patenting over claims 56-126 of copending Application No. 09/524,095. This is a provisional double patenting rejection since the conflicting claims have not yet been patented.

The subject matter claimed in the instant application is fully disclosed in the referenced copending application and would be covered by any patent granted on that copending application since the referenced copending application and the instant application are claiming common subject matter, as follows. Although the conflicting claims are not identical, they are not patentably distinct from each other because it would have been obvious to one of ordinary skill in the art to observed that the omission of the limitations "**soliciting additional input from the user, including user interaction in a modality different that the original request and, refining the navigation query, based upon the additional input**", of applicant claims 56-82 are already in the Co-pending application 09/524,095, as such they are obvious variation of the inventive concept defined in claims 56-126 of the Co-pending application 09/524,095. See In re Karlson, 136USPQ 184 (CCPA 1963). This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented.

### *Claim Rejections - 35 USC § 102*

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent.

5. Claims 56-82 are rejected under 35 U.S.C. 102(e) as being anticipated by Levin et al. (U.S. Patent No. 6,173,279).

6. As per claim 56, Levin et al teach a method for speech-based navigation (*information server, 110*) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (*see abstract, fig*

*1, column 3 lines 5-35*), comprising receiving a spoken request (*receive a natural language query*) for desired information from the user (*user, 112*) utilizing the mobile information appliance (*PC, 102*) of the user; rendering an interpretation (*creating a semantic representation*) of the spoken request, constructing a navigation (*generating search)* query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting (*sending*) the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (*see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22)*

7. As per claim 57, 58, 62-64, Levin et al teach a method of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (*see abstract, fig 1, column 3 lines 5-35*).

8. As per claim 59, Levin et al teach a method of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (*see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22*).

9. As per claim 60, Levin et al teach a method wherein the data link includes a cellular telephone system (*see fig 1, column 2 line 61-67*).

10. As per claim 61, Levin et al teach a method wherein steps (a)-(d) are performed with respect to multiple users (*see abstract, fig 1, column 3 lines 5-35*).

11. As per claim 65, Levin et al teach a computer system for speech-based navigation (*information server, 110*) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (*see abstract, fig 1, column 3 lines 5-35*), comprising a code segment receiving a spoken request (*receive a natural language query*) for desired information from the user (*user*) utilizing the mobile information appliance (*PC, 102*) of the user; a code segment rendering an interpretation (*creating a semantic representation*) of the spoken request, a code segment constructing a navigation (*generating search)* query based upon the interpretation; a code segment utilizing the navigation query to select a portion of the electronic data source; and a code segment transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (*see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22*).

12. As per claim 66, 67, 71-73, Levin et al teach a system of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (*see abstract, fig 1, column 3 lines 5-35*).

13. As per claim 68, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (*see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22*).

14. As per claim 69, Levin et al teach a system wherein the data link includes a cellular telephone system (*see fig 1, column 2 line 61-67*).

15. As per claim 70, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users *(see abstract, fig 1, column 3 lines 5-35*).

16. As per claim 74, Levin et al teach a system for speech-based navigation (*information server, 110*) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (*see abstract, fig 1, column 3 lines 5-35*), comprising receiving a spoken request (*receive a natural language query*) for desired information from the user (*user*) utilizing the mobile information appliance (*PC, 102*) of the user; rendering an interpretation (*creating a semantic representation*) of the spoken request, constructing a navigation (*generating search)* query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting the selected portion of the electronic data source from the network server to the

mobile information appliance of the user. (*see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22*)

17. As per claim 75, 76, 80-81, Levin et al teach a method of rendering the interpretation of the spoken request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (*see abstract, fig 1, column 3 lines 5-35*).

18. As per claim 77, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (*see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22*).

19. As per claim 78, Levin et al teach a system wherein the data link includes a cellular telephone system (*see fig 1, column 2 line 61-67*).

20. As per claim 79, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users (*see abstract, fig 1, column 3 lines 5-35*).

## *Response to Arguments*

21. Applicant's arguments filed on September 26th, 2001 have been fully considered but they are not persuasive.

a. Applicant argues that the prior art "fails to teach or suggest the novel concept of speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where in turn the selected electronic data source from the network server is transmitted to the mobile information appliance of the user." Examiner respectfully disagrees with the applicant perspective and characterization of Levin inventive concept. Levin teach that the URL for a data resource is inputted into PC 102 either by typing the request using a keyboard 104 or ***by speaking the request into a microphone 105,*** which is considered to be a mobile appliance of the user. Furthermore, Levin et al indicate that the spoken requests either from a PC microphone 105 or from a telephone 103 can be handled by a speech recognition system residing at the information server (*see column 4 lines 7-22*). Applicant further argues that the prior art "fails to teach or suggest that the selected electronic data source from the network server is transmitted to the mobile information appliance of the user." Examiner respectfully disagrees with the applicant perspective and characterization of Levin inventive concept. Levin teach that once an information server is accessed, the user can send a text or a spoken query requesting a particular action or service (step 204), for example: "call the pizza place on Main Street in Westfield". The query is received by the access server 106 and the natural language query is sent to the information server 110 via packet network 108. It is to be understood that the packet network 108 may be connected to a plurality of information servers which each relate to one or more particular information services, or there may be a single centralized information server 110 which is accessed by all information services which are capable of receiving and processing natural language queries and contains at least some of the data resources (e.g., URLs and associated site/service-specific grammars) *capable of receiving and responding to a natural language query*. It is obvious inventive concept referring to *response* is in the field of sending or transmitting the requested information to the user. Moreover, it is understood in the art of information request, in order to complete the transaction, the host must transmit to the requester the requested information.

### *Conclusion*

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Firmin Backer whose telephone number is 703-305-0624. The examiner can normally be reached on Mon-Thu 8:30-6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Sheikh Ayaz can be reached on 703-305-9648. The fax phone numbers for the organization where this application or proceeding is assigned are 703-746-7239 for regular communications and 703-746-7238 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

Firmin Backer
Firmin Backer
February 14, 2002

Ayaz Sheikh
**AYAZ SHEIKH**
**SUPERVISORY PATENT EXAMINER**
**TECHNOLOGY CENTER 2100**

**UNITED STATES DEPARTMENT OF COMMERCE**
**Patent and Trademark Office**
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

| SERIAL NUMBER | FILING DATE | FIRST NAMED APPLICANT | ATTORNEY DOCKETT NO. |
|---|---|---|---|
| | | | |

| EXAMINER | |
|---|---|
| | |

| ART UNIT | PAPER NUMBER |
|---|---|
| | 20 |

DATE MAILED:

# EXAMINER INTERVIEW SUMMARY RECORD

All participants (applicant, applicant's representative, PTO personnel):

(1) David Wiley (3) ________

(2) Kin-wah Tong 39,400 (4) ________

Date of Interview 5/23/2002

Type: ☐ Telephonic ☒ Personal (copy is given to ☐ applicant ☐ applicant's representative).

Exhibit shown or demonstration conducted: ☐ Yes ☒ No. If yes, brief description:

Agreement ☒ was reached with respect to some or all of the claims in question. ☐ was not reached.

Claims discussed: 56-82

Identification of prior art discussed: Levin et al.

Description of the general nature of what was agreed to if an agreement was reached, or any other comments: The applicant agreed to amend the claims to further identify the mobile device to overcome the Levin reference.

(A fuller description, if necessary, and a copy of the amendments, if available, which the examiner agreed would render the claims allowable must be attached. Also, where no copy of the amendments which would render the claims allowable is available, a summary thereof must be attached.)

☒ 1. It is not necessary for applicant to provide a separate record of the substance of the interview.

Unless the paragraph below has been checked to indicate to the contrary, A FORMAL WRITTEN RESPONSE TO THE LAST OFFICE ACTION IS NOT WAIVED AND MUST INCLUDE THE SUBSTANCE OF THE INTERVIEW (e.g., items 1-7 on the reverse side of this form). If a response to the last Office action has already been filed, then applicant is given one month from this interview date to provide a statement of the substance of the interview.

☐ 2. Since the examiner's interview summary above (including any attachments) reflects a complete response to each of the objections, rejections and requirements that may be present in the last Office action, and since the claims are now allowable, this completed form is considered to fulfill the response requirements of the last Office action. Applicant is not relieved from providing a separate record of the substance of the interview unless box 1 above is also checked.

Examiner's Signature

PTOL-413 (REV. 2-93)

ORIGINAL FOR INSERTION IN RIGHT HAND FLAP OF FILE WRAPPER

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

| | |
|---|---|
| **PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a)** | Docket Number (Optional)<br>SRI 1P037B |
| In re Application of Halverson, et al | |
| Application Number 09/608,872 | Filed June 30, 2000 |
| For Mobile Navigation of Network-Based Electronic Information Using Spoken Output | |
| Group Art Unit<br>2155 | Examiner<br>F. Backer |

#21
S. Cotton,
7/18/02

This is a request under the provisions of 37 CFR 1.136(a) to extend the period for filing a response in the above identified application.

The requested extension and appropriate non-small-entity fee are as follows (check time period desired):

| | |
|---|---|
| ☐ One month (37 CFR 1.17(a)(1)) | $ |
| ☒ Two months (37 CFR 1.17(a)(2)) | $400.00 |
| ☐ Three months (37 CFR 1.17(a)(3)) | $ |
| ☐ Four months (37 CFR 1.17(a)(4)) | $ |
| ☐ Five months (37 CFR 1.17(a)(5)) | $ |

☒ Applicant claims small entity status. See 37 CFR 1.27. Therefore, the fee amount shown above is reduced by one-half, and the resulting fee is: $ 200.00 .

☐ A check in the amount of the fee is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Commissioner has already been authorized to charge fees in this application to a Deposit Account.

☒ The Commissioner is hereby authorized to charge any fees which may be required, or credit any overpayment, to Deposit Account Number 20-0782 .
I have enclosed a duplicate copy of this sheet.

I am the ☐ applicant/inventor.

☐ assignee of record of the entire interest. See 37 CFR 3.71
Statement under 37 CFR 3.73(b) is enclosed. (Form PTO/SB/96).

☒ attorney or agent of record.

☐ attorney or agent under 37 CFR 1.34(a).
Registration number if acting under 37 CFR 1.34(a). ______ .

**WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

July 17, 2002
Date

Signature
Kin-Wah Tong
Typed or printed name

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required. Submit multiple forms if more than one signature is required, see below*.

☐ *Total of _____ forms are submitted.

Burden Hour Statement: This form is estimated to take 0.1 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

09/608,872

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

## PATENT APPLICATION

#22/D
S. Cotton
7-18-02

Applicant: **Halverson et al.**

Case: **SRI1P037B**

Serial No.: **09/608,872** Filed: **June 30, 2000**

Group Art Unit: **2155**

Examiner: **Firmin Backer**

Title: **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT**

ASSISTANT COMMISSIONER FOR PATENTS
**Box Non-Fee Amendment**
Washington, D. C. 20231

S I R:

## AMENDMENT AND RESPONSE UNDER 37 C.F.R. § 1.111

This amendment addresses the Office Action dated February 19, 2002 (Paper No. 19).

IN THE CLAIMS

**Please amend claims 56, 65 and 74 as shown below. These claims are "clean version" of the amended claims, i.e., with changes incorporated into the claims, whereas the Appendix to this Amendment illustrates the amended claims using underlines and brackets to indicate addition and deletion, respectively.**

1 ~~56.~~ (Twice Amended) A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:

x01

34

D

09/608,872

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

D1

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d) utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

10 ~~85~~. (Twice Amended) A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

D2

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

19 ~~74~~. (Amended) A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

D3

(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises

D

a portable remote control device or a set-top box for a television;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

A03

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user.

## REMARKS

Applicants' representative would like to thank Primary Examiner David Wiley for kindly taking a substantial amount of time on May 23, 2002 to discuss the merits of the subject invention in a face-to-face Examiner Interview. Applicants' representative is aware of the time constraint that is placed on the Examiner and is appreciative of the Examiner's willingness to devote such large quantity of time to discuss the case on the merit.

In view of the following discussion, the Applicants submit that none of the claims now pending in the application are anticipated under the provisions of 35 U.S.C. § 102. Thus, the Applicants believe that all of these claims are now in allowable form.

### I. REJECTION OF CLAIMS 56-82 UNDER DOUBLE PATENTING

The Examiner provisionally rejected claims 56-82 in Paragraphs 2-3 of the Office Action based on the judicially created doctrine of obviousness-type double patenting as being unpatentable over claims 56-126 of copending Application No. 09/524,095.

Responsive to the Examiner, Applicants provisionally agree to file a terminal disclaimer to resolve the present judicially created doctrine of obviousness-type double patenting rejection if and when one of the applications is finally allowed. In accordance with MPEP 804 I.B, "if the 'provisional' double patenting rejection in one application is

36

D

the only rejection remaining in that application, the examiner should then withdraw that rejection and permit the application to issue as a patent, thereby converting the 'provisional' doubling patenting rejection in the other application(s) into a double patenting rejection at the time the one application issues as a patent". As such, Applicants will file a terminal disclaimer in the future, if necessary.

## II. REJECTION OF CLAIMS 56-82 UNDER 35 U.S.C. § 102

The Examiner has again rejected claims 56-82 in Paragraphs 4-20 of the Office Action as being anticipated by the Levin et al. patent (US Patent 6,173,279 issued January 9, 2001, hereinafter referred to as Levin). The rejection is respectfully traversed.

Levin teaches "a method of using at least one natural language query to retrieve information from one or more data resources and further performing a requested action using the retrieved information is disclosed". (See Levin, Column 2, lines 15-18) Namely, Levin teaches a method for using natural language query to obtain information, where upon receipt of the requested information, a desired action is executed based upon the requested information. To illustrate, Levin provides the example, where a user employs natural language to request the telephone number of a restaurant. Upon receipt of the telephone number, the telephone number is actually dialed for the user. (See Levin, Column 3 line 62 to Column 4, line 1)

In contrast, Levin fails to teach or suggest the novel concept of speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television. Specifically, Applicants' independent claims 56, 65 and 74 positively recite:

> 56. A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:
>
> (a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile

D

information appliance comprises a portable remote control device or a set-top box for a television;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d)utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

65. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

74. A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

D

Applicants' invention teaches a novel method and apparatus for speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television. This teaching is completely absent in the Levin reference.

During the Examiner Interview, Primary Examiner David Wiley indicated that a specific identification of the mobile information appliance that comprises a portable remote control device or a set-top box for a television would likely overcome the Levin reference.

Therefore, the Applicants respectfully submit that independent claims 56, 65 and 74 are not anticipated by the Levin reference. As such, claims 56, 65 and 74 fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

Claims 57-64, 66-73 and 75-82 depend, either directly or indirectly, from claims 56, 65 and 74 and recite additional features therefor. Since Levin fails to anticipate Applicants' invention as recited in Applicants' independent claims 56, 65 and 74, dependent claims 57-64, 66-73 and 75-82 are also not anticipated under 35 U.S.C. § 102 and are allowable for the same reason noted above.

## Conclusion

Thus, the Applicants submit that all of these claims now fully satisfy the requirements of 35 U.S.C. §102. Consequently, the Applicants believe that all these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

D

09/608,872

Respectfully submitted,

7/17/02

Kin-Wah Tong, Attorney
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
First Floor,
Shrewsbury, New Jersey 07702

D

## Appendix
## (Marked-up copy of amended claims)

56. (Twice Amended) A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d)utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user[, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication].

65. (Twice Amended) A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) a code segment that renders an interpretation of the spoken request.

(c) a code segment that constructs a navigation query based upon the interpretation;

D

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user[, wherein at least a portion of said data link between said mobile information appliance of the user and the one or more network servers utilizes wireless communication].

74. (Amended) A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user[, wherein at least a portion of a data link of the electronic communications infrastructure between a mobile information appliance of the user and the one or more network servers utilizes wireless communication].

D

# TELEFAX COVER SHEET

**MOSER, PATTERSON & SHERIDAN, LLP**
ATTORNEYS AT LAW
595 SHREWSBURY AVENUE
FIRST FLOOR
SHREWSBURY, NJ 07702
**TELEPHONE (732) 530-9404**
**TELEFAX (732) 530-9808**

**********************************************************

THIS TELEFAX MESSAGE IS ADDRESSED TO THE PERSON OR COMPANY LISTED BELOW. IF IT WAS SENT OR RECEIVED INCORRECTLY, OR YOU ARE NOT THE INTENDED RECIPIENT, PLEASE TAKE NOTICE THAT THIS MESSAGE MAY CONTAIN PRIVILEGED OR CONFIDENTIAL MATERIAL, AND YOUR DUE REGARD FOR THIS INFORMATION IS NECESSARY. YOU MAY ARRANGE TO RETURN THIS MATERIAL BY CALLING THE FIRM LISTED ABOVE AT (732) 530-9404

**********************************************************

THIS MESSAGE HAS 13 PAGES INCLUDING THIS SHEET

TO: Assistant Commissioner of Patents

FAX NO.: 703-746-7239

FROM: Kin-Wah Tong

DATE: July 17, 2002

MATTER: Serial No. 09/608,872 Filed: June 30, 2000

DOCKET NO.: SRI 1P037B

APPLICANT: HALVERSON, et al

The following has been received in the U.S. Patent and Trademark Office on the date of this facsimile:

- ___ Petition
- ___ Disclosure Statement & PTO-1449
- ___ Priority Document
- ___ Drawings (____ sheets) informal
- X Petition for Extension of Time (2 copies)
- X Amendment and Response
- ___ __________
- X Transmittal Letter
- ___ Fee Transmittal (2 copies)
- X Deposit Account Transaction
- X Facsimile Transmission Certificate dated July 17, 2002
- ___ __________
- ___ __________

## CERTIFICATE OF TRANSMISSION UNDER 37 C.F.R. §1.8

I hereby certify that this correspondence is being transmitted by facsimile to the Assistant Commissioner for Patents, Box Non-Fee Amendment, Washington, DC 20231 on July 17, 2002. Facsimile No. 703-746-7239.

Linda DeNardi
Name of person signing this certificate

Linda DeNardi July 17, 2002
Signature and date

Received from < 732 530 9808 > at 7/17/02 5:23:40 PM [Eastern Daylight Time]

D

RECEIVED 7/18/02

**Official**

PTO/SB/21 (08-00)
Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Please type a plus sign (+) inside this box → ☐+

# TRANSMITTAL FORM

*(to be used for all correspondence after initial filing)*

| | |
|---|---|
| **Application Number** | 09/608,872 |
| **Filing Date** | June 30, 2000 |
| **First Named Inventor** | HALVERSON |
| Group Art Unit | 2155 |
| Examiner Name | F. BACKER |
| Attorney Docket Number | SRI 1 P 037B |
| Total Number of Pages in This Submission | 13 |

ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☒ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☒ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Assignment Papers *(for an Application)*
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s)
- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) *(please identify below)*:

  **Certificate of Facsimile Transmission**

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

| | |
|---|---|
| Firm or Individual name | KIN-WAH TONG, ESQ., Reg. No. 39,400 |
| Signature | |
| Date | July 17, 2002 |

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be send to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

D

2155

COPY OF PAPERS ORIGINALLY FILED

| | | | |
|---|---|---|---|
| **TRANSMITTAL OF INF ATION DISCLOSURE STATEM**<br>**Under 37 CFR 1.97(b), (c), or (d)** | | | Docket No.<br>**SRI1P037B** |
| In re Application of: **Halverson, et al.** | | | |
| Serial No.<br>**09/608,872** | Filing Date<br>**June 30, 2000** | Examiner<br>**Firmin Backer** | Group Art Unit<br>**2155** |
| Title: **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT** | | | |

OIPE JC156
JUL 29 2002
PATENT & TRADEMARK OFFICE

Address to:
**Assistant Commissioner for Patents**
**Washington, D.C. 20231**

RECEIVED
JUL 31 2002
Technology Center 2100

## 37 CFR 1.97(b)

1. ☐ The Information Disclosure Statement submitted herewith is being filed within three months of the filing of a national application other than a continued prosecution application under 37 CFR 1.53(d); within three months of the date of entry of the national stage as set forth in 37 CFR 1.491 in an international application; before the mailing of a first Office Action on the merits; or before the mailing of a first Office Action after the filing of a request for continued examination under 37 CFR 1.114.

## 37 CFR 1.97(c)

2. X The Information Disclosure Statement submitted herewith is being filed after the period specified in 37 CFR 1.97(b), but prior to the mailing date of a Final Action under 37 CFR 1.113, a Notice of Allowance under 37 CFR 1.311, or an Action that otherwise closes prosecution in the application, and is accompanied by the statement or fee as indicated below.

## 37 CFR 1.97(d)

3. ☐ The Information Disclosure Statement submitted herewith is being filed after the period specified in 37 CFR 1.97(c), but on or before payment of the issue fee and is accompanied by the statement and fee as indicated below.

## Required Statements and/or Fees Under 37 CFR 1.97(c) or (d)

X Each item of information contained in the accompanying Information Disclosure Statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the Information Disclosure Statement. (37 CFR 1.97(e)(1))

☐ No item of information in the accompanying Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the undersigned person, after making reasonable inquiry, no item of information contained in the accompanying Information Disclosure Statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the Information Disclosure Statement. (37 CFR 1.97(e)(2))

☐ The fee set forth in 37 CFR 1.17(p). Please credit any overpayment or charge any insufficiencies to deposit account number 20-0782.

## 37 CFR §1.704(d)

4. X Each item of information in the accompanying Information Disclosure Statement was cited in a communication from a foreign patent office in a counterpart application and this communication was not received by any individual designated in 37 CFR §1.56(c) more than thirty days prior to the filing of the Information Disclosure Statement.

Dated: July 23 2002

Kin-Wah Tong, Attorney
Reg. No. 39,400

**Moser, Patterson & Sheridan, LLP**
**Attorneys at Law**
**595 Shrewsbury Avenue, Suite 100**
**Shrewsbury, New Jersey 07702**
**732-530-9404**

**Certificate of Mailing by First Class Mail**

I certify that this document is being deposited on **July 23, 2002** with the U.S. Postal Service as first class mail under 37 CFR §1.8 and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231

Barbara Jackson
*Signature of Person Mailing Correspondence*

Barbara J. Jackson
*Typed or Printed Name of Person Mailing Correspondence*

Sheet 1 of 1 sheet(s)

| | | |
|---|---|---|
| U.S. Department of Commerce, Patent and Trademark Office | Docket No. | Serial No. |
| (PTO Form 1449 modified) | SRI1P037B | 09/608,872 |
| **INFORMATION DISCLOSURE STATEMENT BY APPLICANT** | Applicant<br>Halversen, et al. #23 | Confirmation No.:<br>2382 |
| (Use several sheets if necessary) | Filing Date | Group |
| Examiner Firmin Backer | June 30, 2000 | 2155 |

OIPE JC156
JUL 29 2002
PATENT & TRADEMARK OFFICE

COPY OF PAPERS ORIGINALLY FILED

RECEIVED
JUL 31 2002
Technology Center 2100

**U.S. Patent Documents**

| *Examiner Initial | | Document Number | Issue Date | Applicant(s) Name | Class | Subclass | Filing Date If Appropriate |
|---|---|---|---|---|---|---|---|
| FJ | A1 | 6,016,476 | 01/18/2000 | Maes, et al. | 705 | 1 | |
| | A2 | | | | | | |
| | A3 | | | | | | |
| | A4 | | | | | | |
| | A5 | | | | | | |
| | A6 | | | | | | |
| | A7 | | | | | | |
| | A8 | | | | | | |
| | A9 | | | | | | |
| | A10 | | | | | | |
| | A11 | | | | | | |
| | A12 | | | | | | |
| | A13 | | | | | | |

**Foreign Patent Documents**

| *Examiner Initial | | Document Number | Date | Country | Class | Subclass | Translation | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | YES | NO |
| FJ | B1 | 0 867 861 | 09/30/1998 | EPO | G10L | 5/06 | ☐ | ☐ |
| FJ | B2 | 99/50826 | 10/07/1999 | WIPO | G10L | 3/00 | ☐ | ☐ |
| FJ | B3 | 00/05638 | 02/03/2000 | WIPO | G06F | — | ☐ | ☐ |
| | B4 | | | | | | ☐ | ☐ |
| | B5 | | | | | | ☐ | ☐ |

**OTHER ART**

| *Examiner Initial | | Including Author, Title, Date, Pertinent Pages, Etc. |
|---|---|---|
| FJ | C1 | International Search Report, Intl Appl No. PCT/US01/07987 |
| | C2 | |
| | C3 | |

| | |
|---|---|
| Examiner | Date Considered 9/27/02 |

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with your communication to applicant.

# PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:

CARLTON, FIELDS, WARD, EMMANUEL,
SMITH & CUTLER, P.A.
Attn. TONG,Kin-Wah
P.O. Box 3239
TAMPA, FL 33601-3239
UNITED STATES OF AMERICA

# PCT

## NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

Date of mailing (*day/month/year*) 03/07/2002

Applicant's or agent's file reference SRI1P037B.P

**FOR FURTHER ACTION** See paragraphs 1 and 4 below

International application No. PCT/US 01/07987

International filing date (*day/month/year*) 12/03/2001

Applicant

SRI INTERNATIONAL et al.

1. ☒ The applicant is hereby notified that the International Search Report has been established and is transmitted herewith.

**Filing of amendments and statement under Article 19:**
The applicant is entitled, if he so wishes, to amend the claims of the International Application (see Rule 46):

**When?** The time limit for filing such amendments is normally 2 months from the date of transmittal of the International Search Report; however, for more details, see the notes on the accompanying sheet.

**Where?** Directly to the International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland
Fascimile No.: (41-22) 740.14.35

**For more detailed instructions,** see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no International Search Report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. ☐ **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Further action(s):** The applicant is reminded of the following:

Shortly after **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90*bis*.1 and 90*bis*.3, respectively, before the completion of the technical preparations for international publication.

Within **19 months** from the priority date, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later).

Within **20 months** from the priority date, the applicant must perform the prescribed acts for entry into the national phase before all designated Offices which have not been elected in the demand or in a later election within 19 months from the priority date or could not be elected because they are not bound by Chapter II.

| | |
|---|---|
| Name and mailing address of the International Searching Authority<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL-2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Authorized officer<br>Claude Berthon |

Form PCT/ISA/220 (July 1998)

## NOTES TO FORM PCT/ISA/220

These Notes are intended to give the basic instructions concerning the filing of amendments under article 19. The Notes are based on the requirements of the Patent Cooperation Treaty, the Regulations and the Administrative Instructions under that Treaty. In case of discrepancy between these Notes and those requirements, the latter are applicable. For more detailed information, see also the PCT Applicant's Guide, a publication of WIPO.

In these Notes, "Article", "Rule", and "Section" refer to the provisions of the PCT, the PCT Regulations and the PCT Administrative Instructions respectively.

## INSTRUCTIONS CONCERNING AMENDMENTS UNDER ARTICLE 19

The applicant has, after having received the international search report, one opportunity to amend the claims of the international application. It should however be emphasized that, since all parts of the international application (claims, description and drawings) may be amended during the international preliminary examination procedure, there is usually no need to file amendments of the claims under Article 19 except where, e.g. the applicant wants the latter to be published for the purposes of provisional protection or has another reason for amending the claims before international pbulication. Furthermore, it should be emphasized that provisional protection is available in some States only.

**What parts of the international application may be amended?**

Under Article 19, only the claims may be amended.

During the international phase, the claims may also be amended (or further amended) under Article 34 before the International Preliminary Examining Authority. The description and drawings may only be amended under Article 34 before the International Examining Authority.

Upon entry into the national phase, all parts of the international application may be amended under Article 28 or, where applicable, Article 41.

**When?** Within 2 months from the date of transmittal of the international search report or 16 months from the priority date, whichever time limit expires later. It should be noted, however, that the amendments will be considered as having been received on time if they are received by the International Bureau after the expiration of the applicable time limit but before the completion of the technical preparations for international publication (Rule 46.1).

**Where not to file the amendments?**

The amendments may only be filed with the International Bureau and not with the receiving Office or the International Searching Authority (Rule 46.2).

Where a demand for international preliminary examination has been/is filed, see below.

**How?** Either by cancelling one or more entire claims, by adding one or more new claims or by amending the text of one or more of the claims as filed.

A replacement sheet must be submitted for each sheet of the claims which, on account of an amendment or amendments, differs from the sheet originally filed.

All the claims appearing on a replacement sheet must be numbered in Arabic numerals. Where a claim is cancelled, no renumbering of the other claims is required. In all cases where claims are renumbered, they must be renumbered consecutively (Administrative Instructions, Section 205(b)).

**The amendments must be made in the language in which the international application is to be published.**

**What documents must/may accompany the amendments?**

**Letter (Section 205(b)):**

The amendments must be submitted with a letter.

The letter will not be published with the international application and the amended claims. It should not be confused with the "Statement under Article 19(1)" (see below, under "Statement under Article 19(1)").

**The letter must be in English or French, at the choice of the applicant. However, if the language of the international application is English, the letter must be in English; if the language of the international application is French, the letter must be in French.**

## NOTES TO FORM PCT/ISA/220 (continued)

The letter must indicate the differences between the claims as filed and the claims as amended. It must, in particular, indicate, in connection with each claim appearing in the international application (it being understood that identical indications concerning several claims may be grouped),whether

(i) the claim is unchanged;

(ii) the claim is cancelled;

(iii) the claim is new;

(iv) the claim replaces one or more claims as filed;

(v) the claim is the result of the division of a claim as filed.

**The following examples illustrate the manner in which amendments mustbe explained in the accompanying letter:**

1. [Where originally there were 48 claims and after amendment of some claims there are 51]:
"Claims 1 to 29, 31, 32, 34, 35, 37 to 48 replaced by amended claims bearing the same numbers; claims 30, 33 and 36 unchanged; new claims 49 to 51 added."

2. [Where originally there were 15 claims and after amendment of all claims there are 11]:
"Claims 1 to 15 replaced by amended claims 1 to 11."

3. [Where originally there were 14 claims and the amendments consist in cancelling some claims and in adding new claims]:
"Claims 1 to 6 and 14 unchanged; claims 7 to 13 cancelled; new claims 15, 16 and 17 added." or
"Claims 7 to 13 cancelled; new claims 15, 16 and 17 added; all other claims unchanged."

4. [Where various kinds of amendments are made]:
"Claims 1-10 unchanged; claims 11 to 13, 18 and 19 cancelled; claims 14, 15 and 16 replaced by amended claim 14; claim 17 subdivided into amended claims 15, 16 and 17; new claims 20 and 21 added."

**"Statement under article 19(1)" (Rule 46.4)**

The amendments may be accompanied by a statement explaining the amendments and indicating any impact that such amendments might have on the description and the drawings (which cannot be amended under Article 19(1)).

The statement will be published with the international application and the amended claims.

**It must be in the language in which the international appplication is to be published.**

It must be brief, not exceeding 500 words if in English or if translated into English.

It should not be confused with and does not replace the letter indicating the differences between the claims as filed and as amended. It must be filed on a separate sheet and must be identified as such by a heading, preferably by using the words "Statement under Article 19(1)."

It may not contain any disparaging comments on the international search report or the relevance of citations contained in that report. Reference to citations, relevant to a given claim, contained in the international search report may be made only in connection with an amendment of that claim.

**Consequence if a demand for international preliminary examination has already been filed**

If, at the time of filing any amendments under Article 19, a demand for international preliminary examination has already been submitted, the applicant must preferably, at the same time of filing the amendments with the International Bureau, also file a copy of such amendments with the International Preliminary Examining Authority (see Rule 62.2(a), first sentence).

**Consequence with regard to translation of the international application for entry into the national phase**

The applicant's attention is drawn to the fact that, where upon entry into the national phase, a translation of the claims as amended under Article 19 may have to be furnished to the designated/elected Offices, instead of, or in addition to, the translation of the claims as filed.

For further details on the requirements of each designated/elected Office, see Volume II of the PCT Applicant's Guide.

# PATENT COOPERATION TREATY

# PCT

## INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

| | | |
|---|---|---|
| Applicant's or agent's file reference<br>SRI1P037B.P | **FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below. | |
| International application No.<br>PCT/US 01/ 07987 | International filing date *(day/month/year)*<br>12/03/2001 | (Earliest) Priority Date *(day/month/year)*<br>13/03/2000 |
| Applicant<br>SRI INTERNATIONAL et al. | | |

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   ☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   ☐ contained in the international application in written form.

   ☐ filed together with the international application in computer readable form.

   ☐ furnished subsequently to this Authority in written form.

   ☐ furnished subsequently to this Authority in computer readble form.

   ☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   ☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☐ **Certain claims were found unsearchable** (See Box I).

3. ☐ **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

   ☒ the text is approved as submitted by the applicant.

   ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1A

   ☒ as suggested by the applicant.

   ☐ because the applicant failed to suggest a figure.

   ☐ because this figure better characterizes the invention.

   ☐ None of the figures.

Form PCT/ISA/210 (first sheet) (July 1998)

# I ERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/07987

**A. CLASSIFICATION OF SUBJECT MATTER**
IPC 7 H04M3/493 G10L15/22 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04M G10L G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 00 05638 A (MOTOROLA INC)<br>3 February 2000 (2000-02-03)<br>page 4, line 30 -page 5, line 11<br>page 6, line 13 - line 32<br>page 22, line 28 -page 23, line 15<br>figures 3,5A<br>--- | 1-27 |
| A | EP 0 867 861 A (OCTEL COMMUNICATIONS CORP)<br>30 September 1998 (1998-09-30)<br>column 2, line 33 -column 3, line 48<br>--- | 1-27 |
| A | WO 99 50826 A (ANDREA ELECTRONICS CORP ;ANDREA DOUGLAS (US); MARIANO JOSEPH (US))<br>7 October 1999 (1999-10-07)<br>page 3, line 13 - line 17<br>figure 1A<br>---<br>-/-- | 1-27 |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search
26 June 2002

Date of mailing of the international search report
03/07/2002

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL – 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer
Schweitz, M

1

Form PCT/ISA/210 (second sheet) (July 1992)

page 1 of 2

# I‌ .RNATIONAL SEARCH REPORT

International Application No
PCT/US 01/07987

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | US 6 016 476 A (SEDIVY JAN ET AL)<br>18 January 2000 (2000-01-18)<br>column 3, line 17 - line 37<br>----- | 1-27 |

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

page 2 of 2

## II :RNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/07987

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 0005638 A | 03-02-2000 | US 2002006126 A1 | 17-01-2002 |
| | | AU 5006799 A | 14-02-2000 |
| | | AU 5126999 A | 14-02-2000 |
| | | AU 5127099 A | 14-02-2000 |
| | | AU 5227899 A | 14-02-2000 |
| | | CN 1354851 T | 19-06-2002 |
| | | EP 1099152 A1 | 16-05-2001 |
| | | EP 1101343 A1 | 23-05-2001 |
| | | EP 1099146 A2 | 16-05-2001 |
| | | EP 1099213 A1 | 16-05-2001 |
| | | WO 0005861 A1 | 03-02-2000 |
| | | WO 0005708 A1 | 03-02-2000 |
| | | WO 0005643 A1 | 03-02-2000 |
| | | WO 0005638 A2 | 03-02-2000 |
| | | US 6269336 B1 | 31-07-2001 |
| EP 0867861 A | 30-09-1998 | US 6094476 A | 25-07-2000 |
| | | CA 2233019 A1 | 24-09-1998 |
| | | EP 0867861 A2 | 30-09-1998 |
| | | JP 11088502 A | 30-03-1999 |
| | | US 6385304 B1 | 07-05-2002 |
| | | US 6377662 B1 | 23-04-2002 |
| WO 9950826 A | 07-10-1999 | AU 3212899 A | 18-10-1999 |
| | | CA 2323874 A1 | 07-10-1999 |
| | | EP 1066624 A1 | 10-01-2001 |
| | | JP 2002510074 T | 02-04-2002 |
| | | WO 9950826 A1 | 07-10-1999 |
| US 6016476 A | 18-01-2000 | EP 1004099 A1 | 31-05-2000 |
| | | WO 9908238 A1 | 18-02-1999 |
| | | HU 0004470 A2 | 28-05-2001 |
| | | JP 2001512876 T | 28-08-2001 |
| | | PL 338353 A1 | 23-10-2000 |
| | | TW 385400 B | 21-03-2000 |

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
**United States Patent and Trademark Office**
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B | 2382 |

7590 10/04/2002

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER | |
|---|---|
| JEAN, FRANTZ B | |
| **ART UNIT** | **PAPER NUMBER** |
| 2155 | 24 |

DATE MAILED: 10/04/2002

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 07-01)

|  | Application No. | Applicant(s) |  |
|---|---|---|---|
| ***Office Action Summary*** | 09/608,872 | HALVERSEN ET AL. |  |
|  | **Examiner** | **Art Unit** |  |
|  | Frantz B. Jean | 2155 |  |

***-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --***

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *7/29/2002* .

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *56-82* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *56-82* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

**Priority under 35 U.S.C. §§ 119 and 120**

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) *7,23* .
4) ☐ Interview Summary (PTO-413) Paper No(s). _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: .

U.S. Patent and Trademark Office
PTO-326 (Rev. 04-01) **Office Action Summary** Part of Paper No. 24

## DETAILED ACTION

1. This office action is in response to an amendment received on 7/18/02. Claims 56, 65 and 74 were amended. Claims 56-82 are still pending in this application.

### *Information Disclosure Statement*

2. The IDS received on 7/29/02 have been considered.

### *Claim Rejections - 35 USC § 103*

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 56-82 are rejected under 35 U.S.C. 103(a) as being unpatentable over Levin et al. (U.S. Patent No. 6,173,279) in view of Bailey, III US patent No. 6,353,66.

5. As per claim 56, Levin et al teach a method for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising receiving a request (receive a natural language query) for desired information from the user (user, 112) utilizing the mobile appliance (PC, 102) of the user

wherein said mobile information comprises a portable remote control device or top-box for a television; rendering an interpretation (creating a semantic representation) of the request, constructing a navigation (generating search) query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting (sending) the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22). Although Levin teaches natural language, Levin does not explicitly elaborate on a spoken request for desired information from a user. Bailey III is directed to a network and communication access system which includes a spoken (audible) request for desired information from a user (col. 9 lines 47 et seq; col. 3 lines 21 et seq). It would have been obvious to one of ordinary skill in the art at the time of the invention to have combined Bailey's, III features to Levin's because they would have speeded up the communication process while providing a secure system (see Bailey, III col. 4 lines 41 et seq).

6. As per claims 57, 58, 62-64, Levin et al teach a method of rendering the interpretation of the request is performed at the one or more network servers by the mobile information appliance including a Wireless telephone, a portable computer that is a personal digital assistance (See abstract, fig 1, column 3 lines 5-35).

7. As per claim 59, Levin et al teach a method of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

8. As per claim 60, Levin et al teach a method wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

9. As per claim 61, Levin et al teach a method wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

10. As per claim 65, Levin et al teach a computer system for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising a code segment receiving a request (receive a natural language query) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user- a code segment rendering an interpretation (creating a semantic representation) of the request, a code segment constructing a navigation (generating search) query based upon the interpretation; a code segment utilizing the

navigation query to select a portion of the electronic data source; and a code segment transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22). Although Levin teaches natural language, Levin does not explicitly elaborate on a spoken request for desired information from a user. Bailey III is directed to a network and communication access system which includes a spoken (audible) request for desired information from a user (col. 9 lines 47 et seq; col. 3 lines 21 et seq). It would have been obvious to one of ordinary skill in the art at the time of the invention to have combined Bailey's, III features to Levin's because they would have speeded up the communication process while providing a secure system (see Bailey, III col. 4 lines 41 et seq).

11. As per claims 66, 67, 71-73, Levin et al teach a system of rendering the interpretation of the request is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

12. As per claim 68, Levin et at teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion

of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

13. As per claim 69, Levin et al teach a system wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

14. As per claim 70, Levin et a] teach a system wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

15. As per claim 74, Levin et at teach a system for speech-based navigation (information server, 110) of an electronic data source located at one or more network servers located remotely from a user, wherein at least a portion of a data link between a mobile information appliance of the user and the one or more network servers utilizes wireless communication (see abstract, fig 1, column 3 lines 5-35), comprising receiving a request (receive a natural language query) for desired information from the user (user) utilizing the mobile information appliance (PC, 102) of the user; rendering an interpretation (creating a semantic representation) of the request, constructing a navigation (generating search) query based upon the interpretation; utilizing the navigation query to select a portion of the electronic data source; and transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22). Although

Levin teaches natural language, Levin does not explicitly elaborate on a spoken request for desired information from a user. Bailey III is directed to a network and communication access system which includes a spoken (audible) request for desired information from a user (col. 9 lines 47 et seq; col. 3 lines 21 et seq). It would have been obvious to one of ordinary skill in the art at the time of the invention to have combined Bailey's, III features to Levin's because they would have speeded up the communication process while providing a secure system (see Bailey, III col. 4 lines 41 et seq).

16. As per claims 75, 76, 80-8 1, Levin et al teach a method of rendering the interpretation of a request that is performed at the one or more network servers by the mobile information appliance including a wireless telephone, a portable computer that is a personal digital assistance (see abstract, fig 1, column 3 lines 5-35).

17. As per claim 77, Levin et al teach a system of soliciting additional input from the user, including user interaction in a modality different than the original request; refining the navigation query, based upon the additional input; and using the refined navigation query to select a portion of the electronic data source (see abstract, fig. 1-3, column 3 line 36-9 line 5, see also claim 1, 10, 22).

18. As per claim 78, Levin et al teach a system wherein the data link includes a cellular telephone system (see fig 1, column 2 line 61-67).

19. As per claim 79, Levin et al teach a system wherein steps (a)-(d) are performed with respect to multiple users (see abstract, fig 1, column 3 lines 5-35).

### Response to Arguments

20. Applicant's arguments filed on....7/18/02......have been fully considered but they are not persuasive. a. Applicant argues that the prior art "falls to teach or suggest the novel concept of speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user and where in turn the selected electronic data source from the network server is transmitted to the mobile information appliance of the user." Examiner respectfully disagrees with the applicant perspective and characterization of Levin inventive concept. Levin teach that the URL for a data resource is inputted into PC 102 either by typing the request using a keyboard 104 or by speaking the request into a microphone 105, which is considered to be a mobile appliance of the user. Furthermore, Levin et al indicate that the spoken requests either from a PC microphone 105 or from a telephone 103 can be handled by a speech recognition system residing at the information server (see column 4 lines 7-22). Applicant further argues that the prior art "falls to teach or suggest that the selected electronic data source from the network server is transmitted to the mobile information appliance of the user." Examiner respectfully disagrees with the applicant perspective and characterization of Levin inventive concept. Levin teach that once an information server is accessed, the user can

send a text or a spoken query requesting a particular action or service (step 204), for example: "call the pizza place on Main Street in Westfield". The query is received by the access server 106 and the natural language query is sent to the information server I 10 via packet network 108. It is to be understood that the packet network 108 may be connected to a plurality of information servers which each relate to one or more particular information services, or there may be a single centralized information server 110 which is accessed by all information services which are capable of receiving and processing natural language queries and contains at least some of the data resources (e.g., URLs and associated site/service-specific grammars) capable of receiving and responding to a natural language query. It is obvious inventive concept referring to response is in the field of sending or transmitting the requested information to the user. Moreover, it is understood in the art of information request, in order to complete the transaction, the host must transmit to the requester the requested information.

21. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

22. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frantz B. Jean whose telephone number is (703) 305-3970. The examiner can normally be reached on Monday thru Friday from 8:30 to 6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ayaz R. Sheikh, can be reached on (703) 305-9648. The fax phone numbers for this Group are

(703) 746-7238 for After-Final, (703) 746-7239 for Official, and (703) 746-7240 for Non-Official/Draft.

Communications via Internet e-mail regarding this application, other than those under 35 U.S.C. 132 or which otherwise require a signature, may be used by the applicant and should be addressed to [**Ayaz.Sheikh@uspto.gov**].

All Internet e-mail communications will be made of record in the application file. PTO employees do not engage in Internet communications where there exists a possibility that sensitive information could be identified or exchanged unless the record includes a properly signed express waiver of the confidentiality requirements of 35 U.S.C. 122. This is more clearly set forth in the Interim Internet Usage Policy published in the Official Gazette of the Patent and Trademark on February 25, 1997 at 1195 OG 89.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Group receptionist whose telephone number is (703) 305-3900.

Jean

Frantz B. Jean
September 29, 2002
FBJ/

## *Notice of References Cited*

| | | |
|---|---|---|
| Application/Control No.<br>09/608,872 | Applicant(s)/Patent Under Reexamination<br>HALVERSEN ET AL. | |
| Examiner<br>Frantz B. Jean | Art Unit<br>2155 | Page 1 of 1 |

### U.S. PATENT DOCUMENTS

| * | | Document Number<br>Country Code-Number-Kind Code | Date<br>MM-YYYY | Name | Classification |
|---|---|---|---|---|---|
| * | A | US-6,317,684 B1 | 11-2001 | Roeseler et al. | 340/990 |
| * | B | US-6,349,257 B1 | 02-2002 | Liu et al. | 340/5.6 |
| * | C | US-6,314,365 B1 | 11-2001 | Smith, Nicholas E. | 340/988 |
| | D | US-6,353,661 B1 | 03-2002 | Bailey, III, John Edson | 379/88.17 |
| | E | US- | | | |
| | F | US- | | | |
| | G | US- | | | |
| | H | US- | | | |
| | I | US- | | | |
| | J | US- | | | |
| | K | US- | | | |
| | L | US- | | | |
| | M | US- | | | |

### FOREIGN PATENT DOCUMENTS

| * | | Document Number<br>Country Code-Number-Kind Code | Date<br>MM-YYYY | Country | Name | Classification |
|---|---|---|---|---|---|---|
| | N | | | | | |
| | O | | | | | |
| | P | | | | | |
| | Q | | | | | |
| | R | | | | | |
| | S | | | | | |
| | T | | | | | |

### NON-PATENT DOCUMENTS

| * | | Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages) |
|---|---|---|
| | U | |
| | V | |
| | W | |
| | X | |

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

U.S. Patent and Trademark Office
PTO-892 (Rev. 01-2001)

**Notice of References Cited**

Part of Paper No. 24

#25

09/608,872

IN THE UNITED STATES
PATENT AND TRADEMARK OFFICE

## PATENT APPLICATION

Applicant: **Halverson et al.**

Case: **SRI1P037B**

Serial No.: **09/608,872** Filed: **June 30, 2000**

Group Art Unit: **2155**

Examiner: **Frantz Jean**

Title: **MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT**

ASSISTANT COMMISSIONER FOR PATENTS
**Box Non-Fee Amendment**
Washington, D. C. 20231

S I R:

## RESPONSE UNDER 37 C.F.R. § 1.111

This response addresses the Office Action dated October 4, 2002 (Paper No. 24).

## REMARKS

Applicants' representative would like to thank Primary Examiner Frantz Jean for kindly taking a substantial amount of time on December 23, 2002 to discuss the merits of the subject invention in a face-to-face Examiner Interview. Applicants' representative is aware of the time constraint that is placed on the Examiner and is appreciative of the Examiner's willingness to devote such large quantity of time to discuss the case on the merit.

In view of the following discussion, the Applicants submit that none of the claims now pending in the application are made obvious under the provisions of 35 U.S.C. § 103. Thus, the Applicants believe that all of these claims are now in allowable form.

## I. REJECTION OF CLAIMS 56-82 UNDER 35 U.S.C. § 103

The Examiner rejected claims 56-82 in Paragraphs 4-19 of the Office Action as being unpatentable over Levin et al. patent (US Patent 6,173,279 issued January 9, 2001, hereinafter referred to as Levin) in view of Bailey III (US Patent 6,353,661 issued March 5, 2002, hereinafter referred to as Bailey). The rejection is respectfully traversed.

Levin teaches "a method of using at least one natural language query to retrieve information from one or more data resources and further performing a requested action using the retrieved information is disclosed". (See Levin, Column 2, lines 15-18) Namely, Levin teaches a method for using natural language query to obtain information, where upon receipt of the requested information, a desired action is executed based upon the requested information. To illustrate, Levin provides the example, where a user employs natural language to request the telephone number of a restaurant. Upon receipt of the telephone number, the telephone number is actually dialed for the user. (See Levin, Column 3 line 62 to Column 4, line 1)

Bailey teaches a system for using a telephone to interact with a remote system. Specifically, Bailey teaches the use of a conventional phone to allow users to browse, search, store, and create information stored on the Internet. (See Bailey, Abstract; Column 3, lines 8-39)

In contrast, the alleged combination of Levin and Bailey (either singly or in any permissible combination) fails to teach or suggest the novel concept of speech-based navigation <u>where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television</u>. Specifically, Applicants' independent claims 56, 65 and 74 positively recite:

56. A method for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising the steps of:

(a) receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) rendering an interpretation of the spoken request;

(c) constructing a navigation query based upon the interpretation;

(d)utilizing the navigation query to select a portion of the electronic data source; and

(e) transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

65. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, wherein a data link is established between a mobile information appliance of the user and the one or more network servers, comprising:

(a) a code segment that receives a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs a navigation query based upon the interpretation;

(d) a code segment that utilizes the navigation query to select a portion of the electronic data source; and

(e) a code segment that transmits the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

74. A system for speech-based navigation of an electronic data source located at one or more network servers located remotely from a user, comprising:

(a) a mobile information appliance operable to receive a spoken request for desired information from the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television;

(b) spoken language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query based upon the interpretation;

(d) navigation logic, operable to select a portion of the electronic data source using the navigation query, and

(e) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to the mobile information appliance of the user. (emphasis added)

Applicants' invention teaches a novel method and apparatus for speech-based navigation where the method receives spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television. This teaching is completely absent in the Levin and Bailey references.

During the Examiner Interview, Applicants' representative indicated to the Examiner that the present claims specifically recite said mobile information appliance comprises a portable remote control device or a set-top box for a television. Applicants' specification (e.g., on page 2) describes a need for a user interface that does not require the user to learn a highly specialized command language or format. In describing Applicants' invention in the context of a home entertainment setting, Applicants disclose the present invention within the context of a portable remote control device or a set-top box for a television. (e.g., See Applicants' specification, page 6, lines 4-20; and page 18, line 4 to page 19, line 9). In sum, Applicants' novel speech-based navigation method is claimed specifically within the context of a portable remote control device or a set-top box for a television.

During the Examiner Interview, Applicants' representative presented to the Examiner that the combination of Levin and Bailey will fall short of making Applicants' invention obvious. Namely, both references do not disclose Applicants' novel speech-based navigation method within the context of a portable remote control device or a set-top box for a television. For example, Bailey states that "the present invention generally relates to a method and system for combining the power, flexibility, and access to information and communications of the Internet with the simplicity, reliability and wide availability of the existing plain old telephone system (POTS)." (See Bailey, Column 1, lines 5-9) Specifically, the entire purpose of Bailey is to salvage the use of a plain old telephone system to access the Internet. Thus, Bailey does not disclose or suggest Applicants' novel speech-based navigation method within the context of <u>a portable remote control device or a set-top box for a television</u>.

Second, the alleged combination (as taught by Bailey) states that "once the information is obtained the system presents the information to the user by transforming the downloaded text into speech in a manner emulating the behavior of a <u>web browser</u>." (Emphasis added) (See Bailey, Column 3, lines 21-25) Bailey then discloses a complicated method of notifying content, e.g., hyperlinks, of a web page to a user via audible signals. (See Bailey, Column 7, line 5 to Column 8, line 10). In sum, Bailey converts a telephone into a user interface that serves as a web browser as positively asserted by Bailey. This teaching is directly contrary to Applicants' invention which recites "receiving a spoken request for desired information from the user utilizing the mobile information appliance of the user, wherein said mobile information appliance comprises a portable remote control device or a set-top box for a television" and interpreting the spoken request. Applicants' invention is intended to address the criticality of <u>not</u> having to navigate the electronic data source, whereas Bailey simply converts the web page content so that the user is required to manually navigate the data source by listening to different audible signals. Thus, Bailey **<u>teaches away</u>** from Applicants' novel speech-based navigation method.

During the Examiner Interview, the Examiner indicated that he will re-evaluate the cited references and reconsider the present rejections. Therefore, the Applicants respectfully submit that independent claims 56, 65 and 74 are not made obvious by the Levin and Bailey references. As such, claims 56, 65 and 74 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Claims 57-64, 66-73 and 75-82 depend, either directly or indirectly, from claims 56, 65 and 74 and recite additional features therefor. Since Levin and Bailey fail to make Applicants' invention obvious as recited in Applicants' independent claims 56, 65 and 74, dependent claims 57-64, 66-73 and 75-82 are also not made obvious under 35 U.S.C. § 103 and are allowable for the same reason noted above.

## Conclusion

Thus, the Applicants submit that all of these claims now fully satisfy the requirements of 35 U.S.C. §103. Consequently, the Applicants believe that all these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

1/6/03

Kin-Wah Tong, Attorney
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
First Floor,
Shrewsbury, New Jersey 07702

# TELEFAX COVER SHEET

Official
RECEIVED
01-06-03

**MOSER, PATTERSON & SHERIDAN, LLP**
ATTORNEYS AT LAW
595 SHREWSBURY AVENUE
FIRST FLOOR
SHREWSBURY, NJ 07702
**TELEPHONE (732) 530-9404**
**TELEFAX (732) 530-9808**

*********************************************************

THIS TELEFAX MESSAGE IS ADDRESSED TO THE PERSON OR COMPANY LISTED BELOW. IF IT WAS SENT OR RECEIVED INCORRECTLY, OR YOU ARE NOT THE INTENDED RECIPIENT, PLEASE TAKE NOTICE THAT THIS MESSAGE MAY CONTAIN PRIVILEGED OR CONFIDENTIAL MATERIAL, AND YOUR DUE REGARD FOR THIS INFORMATION IS NECESSARY. YOU MAY ARRANGE TO RETURN THIS MATERIAL BY CALLING THE FIRM LISTED ABOVE AT (732) 530-9404

*********************************************************

THIS MESSAGE HAS 9 PAGES INCLUDING THIS SHEET

TO: Commissioner of Patents

FAX NO.: 703-746-7239

FROM: Kin-Wah Tong

DATE: January 6, 2003

MATTER: Serial No. 09/608,872 Filed: June 30, 2000

DOCKET NO.: SRI 1P037B

APPLICANT: HALVERSON, et al

The following has been received in the U.S. Patent and Trademark Office on the date of this facsimile:

___ Petition
___ Disclosure Statement & PTO-1449
___ Priority Document
___ Drawings (____ sheets) informal
___ Petition for Extension of Time (2 copies)
X Response
___ ____________________

X Transmittal Letter (2 copies)
___ Fee Transmittal (2 copies)
___ Deposit Account Transaction
X Facsimile Transmission Certificate dated January 6, 2003
___ ____________________
___ ____________________

## CERTIFICATE OF TRANSMISSION UNDER 37 C.F.R. §1.8

I hereby certify that this correspondence is being transmitted by facsimile to the Commissioner for Patents, Box Non-Fee Amendment, Washington, DC 20231 on **January 6, 2003**, Facsimile No. 703-746-7239.

Kin-Wah Tong
Name of person signing this certificate

January 6, 2003
Signature and date

Received from < 732 530 9808 > at 1/6/03 7:03:49 PM [Eastern Standard Time]

PTO/SB/21 (08-00)

Please type a plus sign (+) inside this box ⟶ [+]

Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

# TRANSMITTAL FORM

*(to be used for all correspondence after initial filing)*

| | |
|---|---|
| **Application Number** | 09/608,872 |
| **Filing Date** | June 30, 2000 |
| **First Named Inventor** | HALVERSON |
| Group Art Unit | 2155 |
| Examiner Name | FRANTZ JEAN |
| Attorney Docket Number | SRI 1 P 037B |

Total Number of Pages in This Submission

ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☒ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53

- ☐ Assignment Papers *(for an Application)*
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____

- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- ☐ Proprietary Information
- ☐ Status Letter
- ☒ Other Enclosure(s) *(please identify below)*:
  **Certificate of Facsimile Transmission**

Remarks: **It is believed no fee is due. However, in the event a fee is due, kindly charge that fee to deposit account number 20-0782. To facilitate that charge, a duplicate copy of this letter is enclosed**

## SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

| | |
|---|---|
| Firm or Individual name | KIN-WAH TONG, ESQ., Reg. No. 39,400 |
| Signature | |
| Date | January 6, 2003 |

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be send to the Chief Information Officer, U.S. Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

PTO/SB/21 (08-00)

Please type a plus sign (+) inside this box ⟶ [+]

Approved for use through 10/31/2002. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

# TRANSMITTAL FORM

*(to be used for all correspondence after initial filing)*

| | |
|---|---|
| **Application Number** | 09/608,872 |
| **Filing Date** | June 30, 2000 |
| **First Named Inventor** | HALVERSON |
| Group Art Unit | 2155 |
| Examiner Name | FRANTZ JEAN |
| Attorney Docket Number | SRI 1 P 037B |

Total Number of Pages in This Submission

ENCLOSURES *(check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☒ Amendment / Response
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Response to Missing Parts/ Incomplete Application
  - ☐ Response to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Assignment Papers *(for an Application)*
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
- ☐ After Allowance Communication to Group
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☐ Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- ☐ Proprietary Information
- ☐ Status Letter
- ☒ Other Enclosure(s) *(please identify below)*:

  **Certificate of Facsimile Transmission**

Remarks: **It is believed no fee is due. However, in the event a fee is due, kindly charge that fee to deposit account number 20-0782. To facilitate that charge, a duplicate copy of this letter is enclosed**

## SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

| | |
|---|---|
| Firm or Individual name | KIN-WAH TONG, ESQ., Reg. No. 39,400 |
| Signature | |
| Date | January 6, 2003 |

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRIlp037B | 2382 |

7590 01/09/2003

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER | |
|---|---|
| JEAN, FRANTZ B | |
| ART UNIT | PAPER NUMBER 26 |
| 2155 | |

DATE MAILED: 01/09/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 07-01)

| | | | |
|---|---|---|---|
| ***Interview Summary*** | **Application No.**<br>09/608,872 | **Applicant(s)**<br>HALVERSEN ET AL. | |
| | **Examiner**<br>Frantz B. Jean | **Art Unit**<br>2155 | |

All participants (applicant, applicant's representative, PTO personnel):

(1) *Frantz B. Jean*. (3)_____.

(2) *Kin-Wah Tong*. (4)_____.

Date of Interview: *23 December 2002* .

Type: a)☐ Telephonic b)☐ Video Conference
c)☒ Personal [copy given to: 1)☐ applicant 2)☒ applicant's representative]

Exhibit shown or demonstration conducted: d)☐ Yes e)☒ No.
If Yes, brief description: _____ .

Claim(s) discussed: Independent claims .

Identification of prior art discussed: Levine & Bailey .

Agreement with respect to the claims f)☐ was reached. g)☒ was not reached. h)☐ N/A.

Substance of Interview including description of the general nature of what was agreed to if an agreement was reached, or any other comments: See below .

(A fuller description, if necessary, and a copy of the amendments which the examiner agreed would render the claims allowable, if available, must be attached. Also, where no copy of the amendments that would render the claims allowable is available, a summary thereof must be attached.)

i)☑ It is not necessary for applicant to provide a separate record of the substance of the interview(if box is checked).

Unless the paragraph above has been checked, THE FORMAL WRITTEN REPLY TO THE LAST OFFICE ACTION MUST INCLUDE THE SUBSTANCE OF THE INTERVIEW. (See MPEP Section 713.04). If a reply to the last Office action has already been filed, APPLICANT IS GIVEN ONE MONTH FROM THIS INTERVIEW DATE TO FILE A STATEMENT OF THE SUBSTANCE OF THE INTERVIEW. See Summary of Record of Interview requirements on reverse side or on attached sheet.

Applicants' representative believes that the invention as claimed does define over the prior art of record Levine & Bailey. Examiner disagrees. Examiner has decided to review & go through Levine & Bailey prior art for further consideration.

Examiner Note: You must sign this form unless it is an Attachment to a signed Office action.

Examiner's signature, if required

U.S. Patent and Trademark Office
PTO-413 (Rev. 03- 98) **Interview Summary** Paper No. 25

| | | |
|---|---|---|
| ***Notice of Allowability*** | **Application No.**<br>09/608,872 | **Applicant(s)**<br>HALVERSEN ET AL. |
| | **Examiner**<br>Frantz B. Jean | **Art Unit**<br>2155 |

***-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--***

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to _the response filed on 1/06/2003_.
2. ☒ The allowed claim(s) is/are _56-82_.
3. ☐ The drawings filed on _____ are accepted by the Examiner.
4. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
   a) ☐ All b) ☐ Some* c) ☐ None of the:
      1. ☐ Certified copies of the priority documents have been received.
      2. ☐ Certified copies of the priority documents have been received in Application No. _____ .
      3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).
   * Certified copies not received: _____.
5. ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
   (a) ☐ The translation of the foreign language provisional application has been received.
6. ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application. **THIS THREE-MONTH PERIOD IS NOT EXTENDABLE**

7. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.

8. ☒ CORRECTED DRAWINGS must be submitted.
   (a) ☒ including changes required by the Notice of Draftsperson's Patent Drawing Review ( PTO-948) attached
      1) ☒ hereto or 2) ☐ to Paper No. ____.
   (b) ☐ including changes required by the proposed drawing correction filed _____, which has been approved by the Examiner.
   (c) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No. _____.

**Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the top margin (not the back) of each sheet. The drawings should be filed as a separate paper with a transmittal letter addressed to the Official Draftsperson.**

9. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

**Attachment(s)**

1☐ Notice of References Cited (PTO-892)
3☒ Notice of Draftperson's Patent Drawing Review (PTO-948)
5☐ Information Disclosure Statements (PTO-1449), Paper No. ____.
7☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material

2☐ Notice of Informal Patent Application (PTO-152)
4☐ Interview Summary (PTO-413), Paper No.____ .
6☐ Examiner's Amendment/Comment
8☒ Examiner's Statement of Reasons for Allowance
9☐ Other

U.S. Patent and Trademark Office
PTO-37 (Rev. 04-01)

**Notice of Allowability**

Part of Paper No. 27

1. Claims 56-82 are allowable over the prior art made of record and in light of Applicants' arguments..

2. The response filed on 01/08/2003 has been entered.

## Reasons for Allowance

3. The examiner respectfully submits that the specific techniques of providing a speech-based navigation where a spoken request for desired information is received from a user utilizing a mobile information appliance of the user, wherein the mobile information appliance comprises a portable remote control device or a set-top box for a television; in conjunction with the other limitations of the dependent and independent claims 56-82 were not shown by, would not have been obvious over, nor would have been fairly suggested by the prior art made of record.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

4. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frantz B. Jean whose telephone number is (703) 305-3970. The examiner can normally be reached on Monday thru Friday from 8:30 to 6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ayaz R. Sheikh, can be reached on (703) 305-9648. The fax phone numbers for this Group are (703) 746-7238 for After-Final, (703) 746-7239 for Official, and (703) 746-7240 for Non-Official/Draft.

Communications via Internet e-mail regarding this application, other than those under 35 U.S.C. 132 or which otherwise require a signature, may be used by the applicant and should be addressed to [Ayaz.Sheikh@uspto.gov].

All Internet e-mail communications will be made of record in the application file. PTO employees do not engage in Internet communications where there exists a possibility that sensitive information could be identified or exchanged unless the record includes a properly signed express waiver of the confidentiality requirements of 35 U.S.C. 122. This is more clearly set forth in the Interim Internet Usage Policy published in the Official Gazette of the Patent and Trademark on February 25, 1997 at 1195 OG 89.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Group receptionist whose telephone number is (703) 305-3900.

Frantz B. Jean
March 07, 2003
FBJ/

Form PTO 948 (Rev. 03/01) U.S. DEPARTMENT OF COMMERCE - Patent and Trademark Office Application No. 09/608872

# NOTICE OF DRAFTSPERSON'S PATENT DRAWING REVIEW

The drawing(s) filed (insert date) 6-30-00 are:

A. ☐ approved by the Draftsperson under 37 CFR 1.84 or 1.152.

B. ☒ objected to by the Draftsperson under 37 CFR 1.84 or 1.152 for the reasons indicated below. The Examiner will require submission of new, corrected drawings when necessary. Corrected drawing must be sumitted according to the instructions on the back of this notice.

1. DRAWINGS. 37 CFR 1.84(a): Acceptable categories of drawings: Black ink. Color.
   - ____ Color drawings are not acceptable until petiton is granted. Fig(s) ____
   - ____ Pencil and non black ink not permitted. Fig(s) ____
2. PHOTOGRAPHS. 37 CFR 1.84(b)
   - ____ 1 full-tone set is required. Fig(s) ____
   - ____ Photographs may not be mounted. 37 CFR 1.84(e)
   - ____ Poor quality (half-tone). Fig(s) ____
3. TYPE OF PAPER. 37 CFR 1.84(e)
   - ____ Paper not flexible, strong, white, and durable. Fig(s) ____
   - ____ Erasures, alterations, overwritings, interlineations, folds, copy machine marks not accepted. Fig(s) ____
   - ____ Mylar, velum paper is not acceptable (too thin). Fig(s) ____
4. SIZE OF PAPER. 37 CFR 1.84(f): Acceptable sizes:
   - ____ 21.0 cm by 29.7 cm (DIN size A4)
   - ____ 21.6 cm by 27.9 cm (8 1/2 x 11 inches)
   - ____ All drawing sheets not the same size. Sheet(s) ____
   - ____ Drawings sheets not an acceptable size. Fig(s) ____
5. MARGINS. 37 CFR 1.84(g): Acceptable margins:

   Top 2.5 cm Left 2.5cm Right 1.5 cm Bottom 1.0 cm
   SIZE: A4 Size
   Top 2.5 cm Left 2.5 cm Right 1.5 cm Bottom 1.0 cm
   SIZE: 8 1/2 x 11
   Margins not acceptable. Fig(s) ____
   ____ Top (T) ____ Left (L)
   ____ Right (R) ____ Bottom (B)
6. VIEWS. 37 CFR 1.84(h)
   REMINDER: Specification may require revision to correspond to drawing changes.
   Partial views. 37 CFR 1.84(h)(2)
   - ____ Brackets needed to show figure as one entity. Fig(s) ____
   - ____ Views not labeled separately or properly. Fig(s) ____
   - ____ Enlarged view not labeled separetely or properly. Fig(s) ____
7. SECTIONAL VIEWS. 37 CFR 1.84 (h)(3)
   - ____ Hatching not indicated for sectional portions of an object. Fig(s) ____
   - ____ Sectional designation should be noted with Arabic or Roman numbers. Fig(s) ____
8. ARRANGEMENT OF VIEWS. 37 CFR 1.84(i)
   - ____ Words do not appear on a horizontal, left-to-right fashion when page is either upright or turned so that the top becomes the right side, except for graphs. Fig(s) ____
9. SCALE. 37 CFR 1.84(k)
   - ____ Scale not large enough to show mechanism without crowding when drawing is reduced in size to two-thirds in reproduction. Fig(s) ____
10. CHARACTER OF LINES, NUMBERS, & LETTERS. 37 CFR 1.84(i)
    - X Lines, numbers & letters not uniformly thick and well defined, clean, durable, and black (poor line quality). Fig(s) ALL
11. SHADING. 37 CFR 1.84(m)
    - ____ Solid black areas pale. Fig(s) ____
    - X Solid black shading not permitted. Fig(s) 6
    - ____ Shade lines, pale, rough and blurred. Fig(s) ____
12. NUMBERS, LETTERS, & REFERENCE CHARACTERS. 37 CFR 1.84(p)
    - X Numbers and reference characters not plain and legible. Fig(s) ALL
    - X Figure legends are poor. Fig(s) ALL
    - ____ Numbers and reference characters not oriented in the same direction as the view. 37 CFR 1.84(p)(1) Fig(s) ____
    - ____ English alphabet not used. 37 CFR 1.84(p)(2) Figs ____
    - ____ Numbers, letters and reference characters must be at least .32 cm (1/8 inch) in height. 37 CFR 1.84(p)(3) Fig(s) ____
13. LEAD LINES. 37 CFR 1.84(q)
    - ____ Lead lines cross each other. Fig(s) ____
    - ____ Lead lines missing. Fig(s) ____
14. NUMBERING OF SHEETS OF DRAWINGS. 37 CFR 1.84(t)
    - ____ Sheets not numbered consecutively, and in Arabic numerals beginning with number 1. Sheet(s) ____
15. NUMBERING OF VIEWS. 37 CFR 1.84(u)
    - ____ Views not numbered consecutively, and in Arabic numerals, beginning with number 1. Fig(s) ____
16. CORRECTIONS. 37 CFR 1.84(w)
    - ____ Corrections not made from prior PTO-948 dated ____
17. DESIGN DRAWINGS. 37 CFR 1.152
    - ____ Surface shading shown not appropriate. Fig(s) ____
    - ____ Solid black shading not used for color contrast. Fig(s) ____

COMMENTS

REVIEWER ____ DATE 3-10-03 TELEPHONE NO. ____

ATTACHMENT TO PAPER NO. 27

&

UNITED STATES DEPARTMENT OF COMMERCE
**United States Patent and Trademark Office**
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

#27

# NOTICE OF ALLOWANCE AND FEE(S) DUE

7590 03/11/2003

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER | |
|---|---|
| JEAN, FRANTZ B | |
| **ART UNIT** | **CLASS-SUBCLASS** |
| 2155 | 709-218000 |

DATE MAILED: 03/11/2003

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRILP037B | 2382 |

TITLE OF INVENTION: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

| APPLN. TYPE | SMALL ENTITY | ISSUE FEE | PUBLICATION FEE | TOTAL FEE(S) DUE | DATE DUE |
|---|---|---|---|---|---|
| nonprovisional | YES | $650 | $0 | $650 | 06/11/2003 |

**THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.**

**THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE REFLECTS A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE APPLIED IN THIS APPLICATION. THE PTOL-85B (OR AN EQUIVALENT) MUST BE RETURNED WITHIN THIS PERIOD EVEN IF NO FEE IS DUE OR THE APPLICATION WILL BE REGARDED AS ABANDONED.**

**HOW TO REPLY TO THIS NOTICE:**

I. Review the SMALL ENTITY status shown above.

If the SMALL ENTITY is shown as YES, verify your current SMALL ENTITY status:

A. If the status is the same, pay the TOTAL FEE(S) DUE shown above.

B. If the status is changed, pay the PUBLICATION FEE (if required) and twice the amount of the ISSUE FEE shown above and notify the United States Patent and Trademark Office of the change in status, or

If the SMALL ENTITY is shown as NO:

A. Pay TOTAL FEE(S) DUE shown above, or

B. If applicant claimed SMALL ENTITY status before, or is now claiming SMALL ENTITY status, check the box below and enclose the PUBLICATION FEE and 1/2 the ISSUE FEE shown above.

☐ Applicant claims SMALL ENTITY status. See 37 CFR 1.27.

II. PART B - FEE(S) TRANSMITTAL should be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). Even if the fee(s) have already been paid, Part B - Fee(s) Transmittal should be completed and returned. If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Box ISSUE FEE unless advised to the contrary.

**IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.**

PTOL-85 (REV. 04-02) Approved for use through 01/31/2004.

# PART B - FEE(S) TRANSMITTAL

**Complete and send this form, together with applicable fee(s), to: Mail Box ISSUE FEE**
**Commissioner for Patents**
**Washington, D.C. 20231**
**Fax (703)746-4000**

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 4 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Legibly mark-up with any corrections or use Block 1)

7590 03/11/2003

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

**Certificate of Mailing or Transmission**

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Box Issue Fee address above, or being facsimile transmitted to the USPTO, on the date indicated below.

(Depositor's name)

(Signature)

(Date)

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRILP037B | 2382 |

TITLE OF INVENTION: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

| APPLN. TYPE | SMALL ENTITY | ISSUE FEE | PUBLICATION FEE | TOTAL FEE(S) DUE | DATE DUE |
|---|---|---|---|---|---|
| nonprovisional | YES | $650 | $0 | $650 | 06/11/2003 |

| EXAMINER | ART UNIT | CLASS-SUBCLASS |
|---|---|---|
| JEAN, FRANTZ B | 2155 | 709-218000 |

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1________________

2________________

3________________

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been previously submitted to the USPTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE

(B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) ☐ individual ☐ corporation or other private group entity ☐ government

4a. The following fee(s) are enclosed:

☐ Issue Fee

☐ Publication Fee

☐ Advance Order - # of Copies ____________

4b. Payment of Fee(s):

☐ A check in the amount of the fee(s) is enclosed.

☐ Payment by credit card. Form PTO-2038 is attached.

☐ The Commissioner is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number ____________(enclose an extra copy of this form).

Commissioner for Patents is requested to apply the Issue Fee and Publication Fee (if any) or to re-apply any previously paid issue fee to the application identified above.

(Authorized Signature) (Date)

NOTE; The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, Washington, DC 20231.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMIT THIS FORM WITH FEE(S)

PTOL-85 (REV. 04-02) Approved for use through 01/31/2004. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRILP037B | 2382 |

7590 03/11/2003

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

| EXAMINER | |
|---|---|
| JEAN, FRANTZ B | |
| **ART UNIT** | **PAPER NUMBER** |
| 2155 | |

DATE MAILED: 03/11/2003

## Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)

**(application filed on or after May 29, 2000)**

The patent term adjustment to date is 0 days. If the issue fee is paid on the date that is three months after the mailing date of this notice and the patent issues on the Tuesday before the date that is 28 weeks (six and a half months) after the mailing date of this notice, the term adjustment will be 0 days.

If a continued prosecution application (CPA) was filed in the above-identified application, the filing date that determines patent term adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) system. (http://pair.uspto.gov)

Any questions regarding the patent term extension or adjustment determination should be directed to the Office of Patent Legal Administration at (703)305-1383.

PTOL-85 (REV. 04-02) Approved for use through 01/31/2004.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRILP037B | 2382 |

7590 03/11/2003

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702
UNITED STATES

| EXAMINER | |
|---|---|
| JEAN, FRANTZ B | |
| ART UNIT | PAPER NUMBER |
| 2155 | |

DATE MAILED: 03/11/2003

## Notice of Fee Increase on January 1, 2003

If a reply to a "Notice of Allowance and Fee(s) Due" is filed in the Office on or after January 1, 2003, then the amount due will be higher than that set forth in the "Notice of Allowance and Fee(s) Due" since there will be an increase in fees effective on January 1, 2003. See Revision of Patent and Trademark Fees for Fiscal Year 2003; Final Rule, 67 Fed. Reg. 70847, 70849 (November 27, 2002).

The current fee schedule is accessible from: http://www.uspto.gov/main/howtofees.htm.

If the issue fee paid is the amount shown on the "Notice of Allowance and Fee(s) Due," but not the correct amount in view of the fee increase, a "Notice to Pay Balance of Issue Fee" will be mailed to applicant. In order to avoid processing delays associated with mailing of a "Notice to Pay Balance of Issue Fee," if the response to the Notice of Allowance and Fee(s) due form is to be filed on or after January 1, 2003 (or mailed with a certificate of mailing on or after January 1, 2003), the issue fee paid should be the fee that is required at the time the fee is paid. If the issue fee was previously paid, and the response to the "Notice of Allowance and Fee(s) Due" includes a request to apply a previously-paid issue fee to the issue fee now due, then the difference between the issue fee amount at the time the response is filed and the previously paid issue fee should be paid. See Manual of Patent Examining Procedure, Section 1308.01 (Eighth Edition, August 2001).

Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at (703) 305-8283.

PTOL-85 (REV. 04-02) Approved for use through 01/31/2004.

#23
JC

OIPE JC54
MAY 06 2003
PATENT & TRADEMARK OFFICE

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

|  |  |  |
|---|---|---|
| In re application of: | Halverson, et al. |  |
| Serial No.: | 09/608,872 | Art Unit: 2155 |
| Filing Date: | June 30, 2000 | Examiner: Jean, Frantz B |
| For: | MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT |  |
| Docket No. | SRI 4116-6 |  |

Assistant Commissioner for Patents
Washington, D.C. 20231
S I R:

## SUBMISSION OF FORMAL DRAWINGS

The Applicants submit herewith 7 sheets of formal drawings (FIGS. 1 through 6), properly labeled, in connection with the above-captioned application. The Examiner is requested to substitute these formal drawings for the informal drawings previously submitted.

Respectfully submitted,

Dated: 4/29/03

KIN-WAH TONG
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
Suite 100
Shrewsbury, NJ 07702

CERTIFICATE OF MAILING under 37 C.F.R. 1.8(a)

I hereby certify that this correspondence is being deposited on April 30, 2003, with the United States Postal Service as first class mail, with sufficient postage, in an envelope addressed to the Commissioner for Patents, Box Issue Fee, Washington, D.C. 20231.

Signature

April 30, 2003
Date of signature

6757718

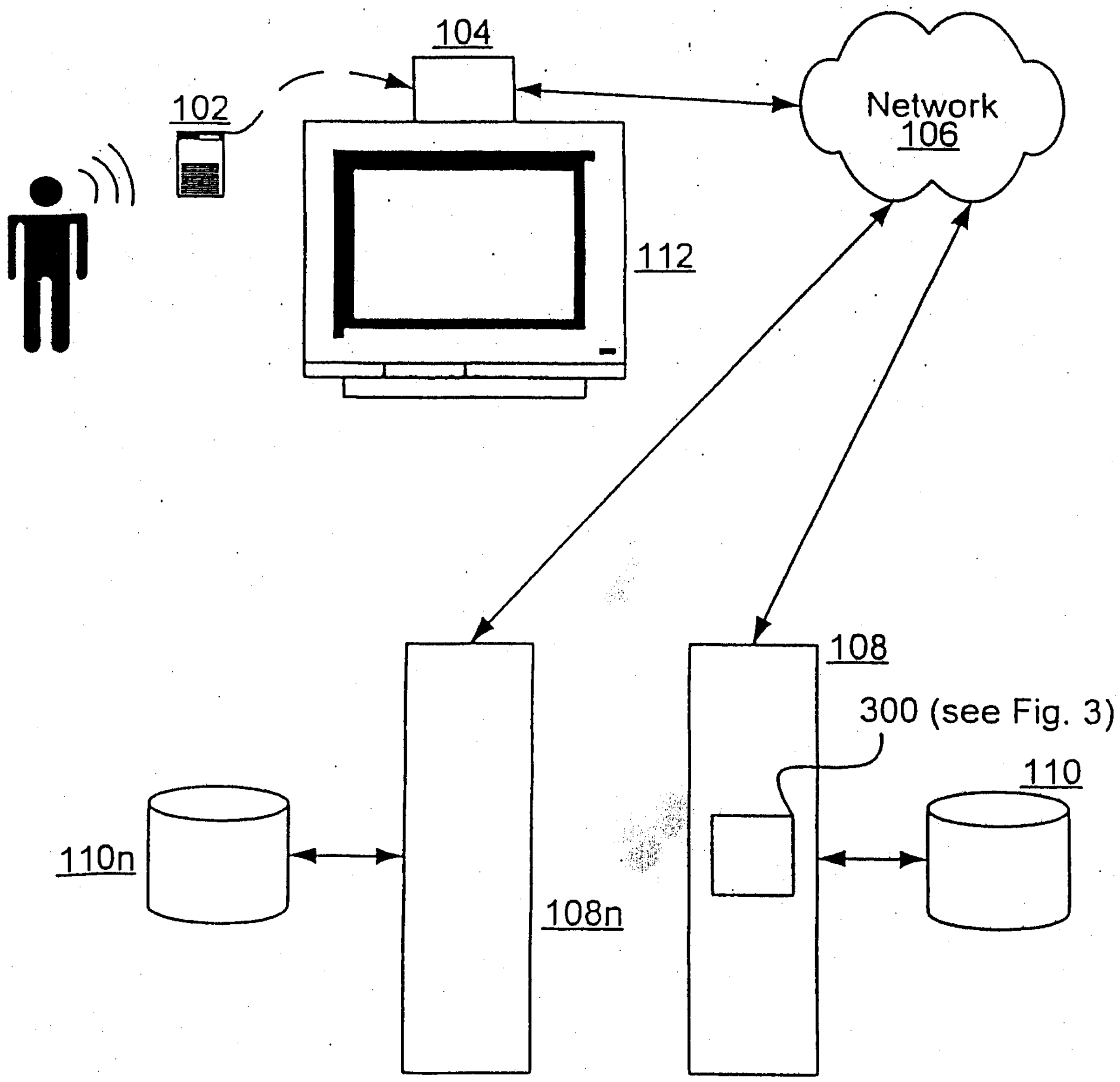


Fig. 1a

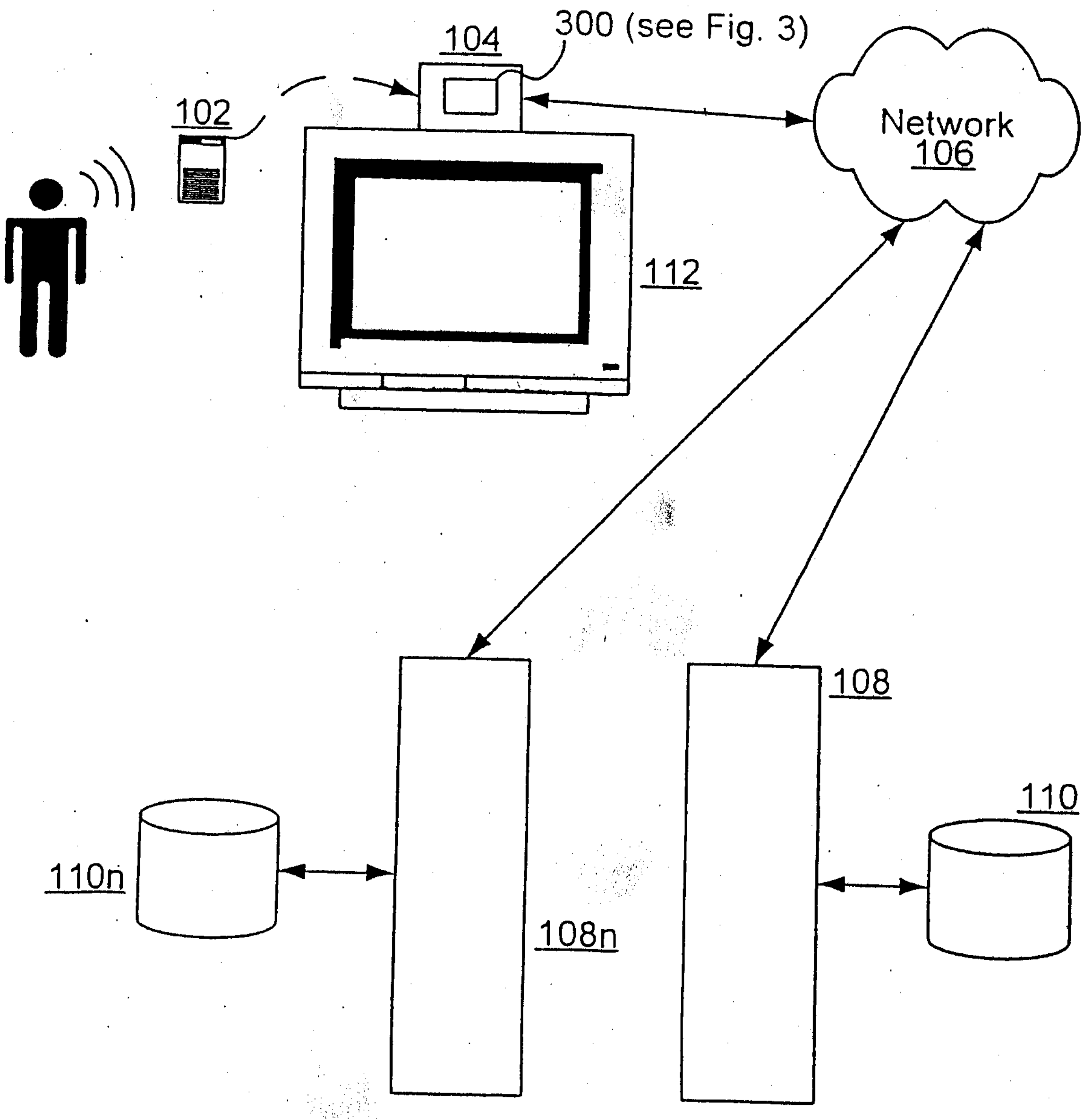


Fig. 1b

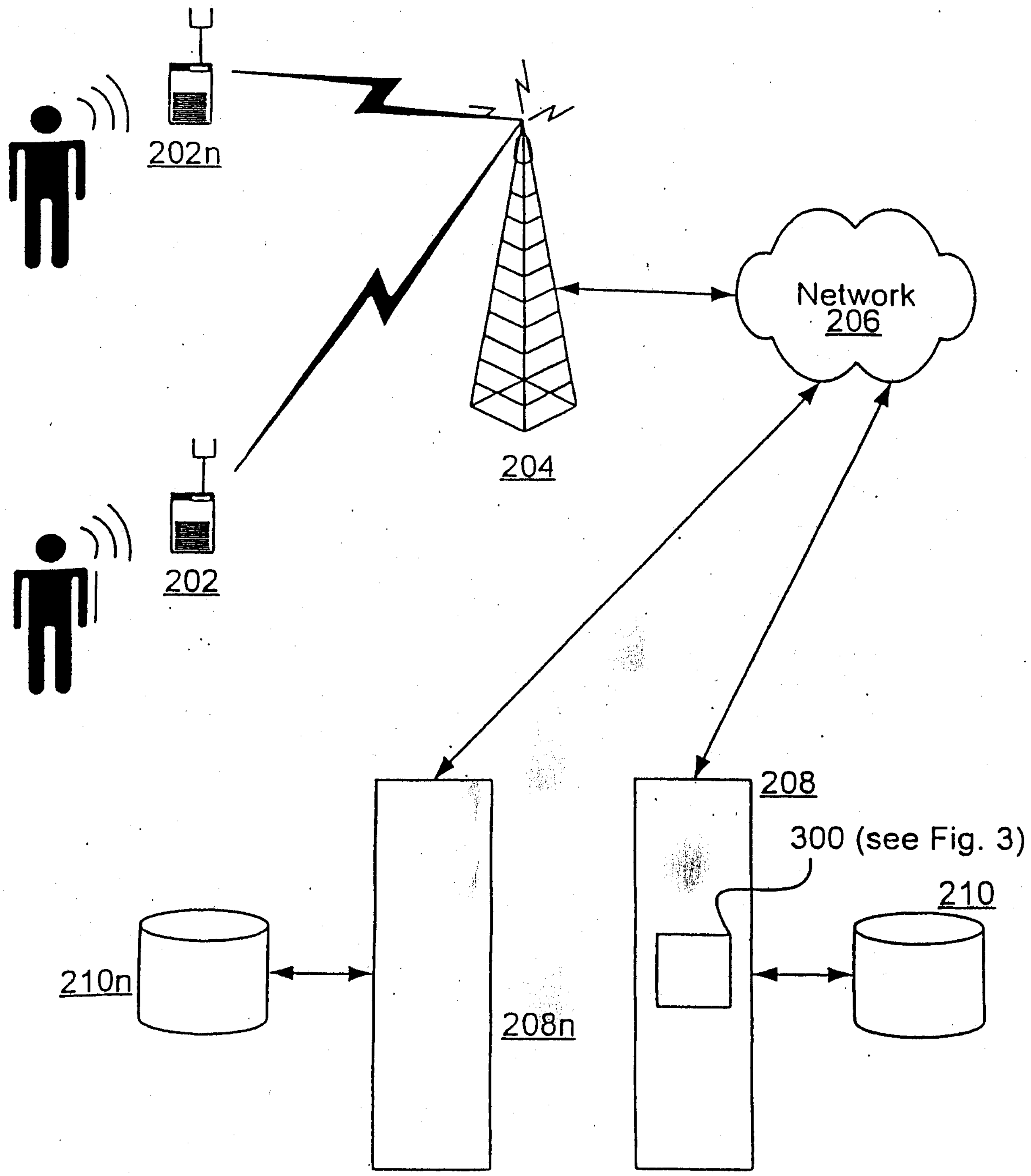


Fig. 2

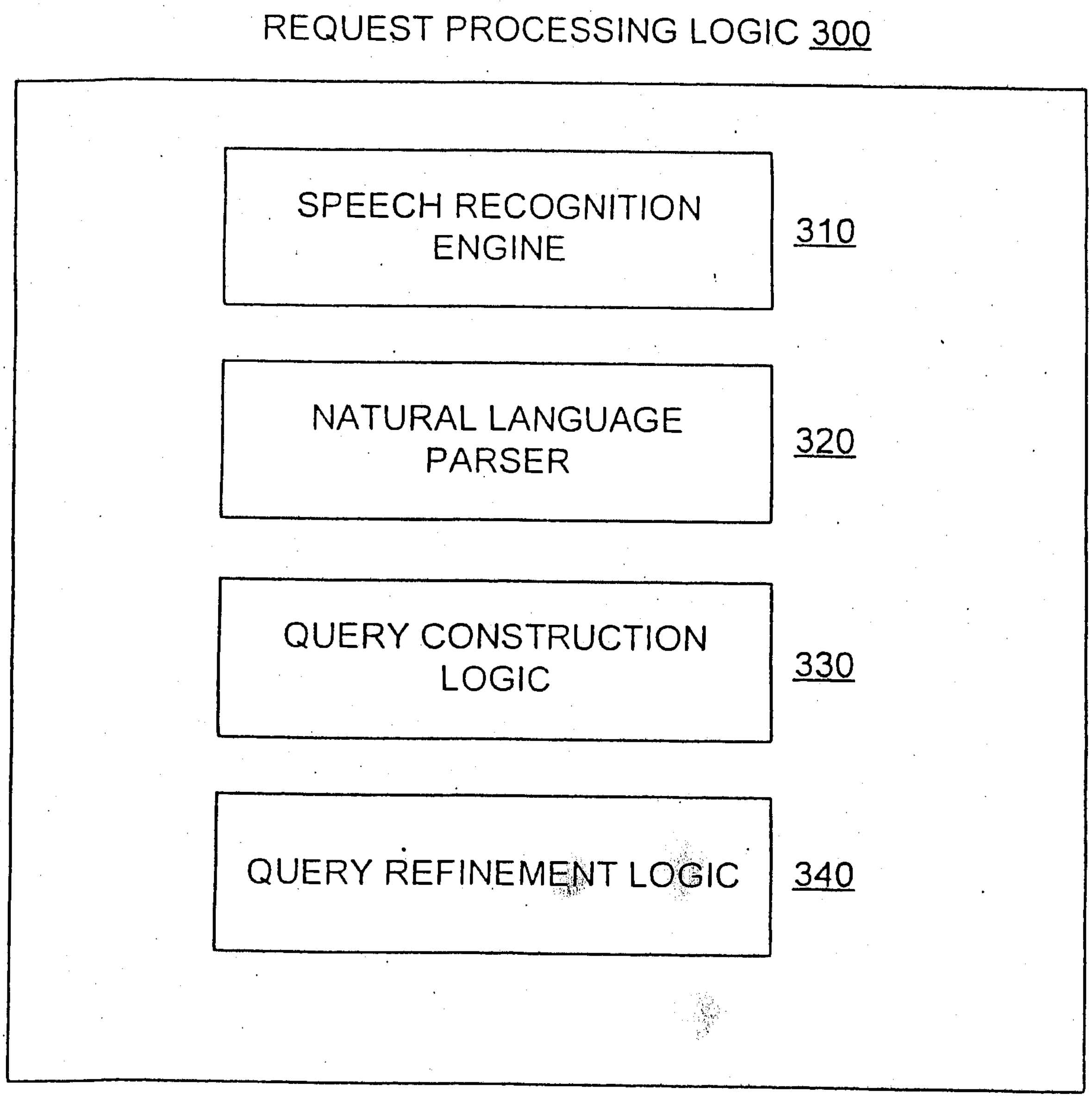

REQUEST PROCESSING LOGIC 300
SPEECH RECOGNITION ENGINE
310
NATURAL LANGUAGE PARSER
320
QUERY CONSTRUCTION LOGIC
330
QUERY REFINEMENT LOGIC
340

Fig. 3

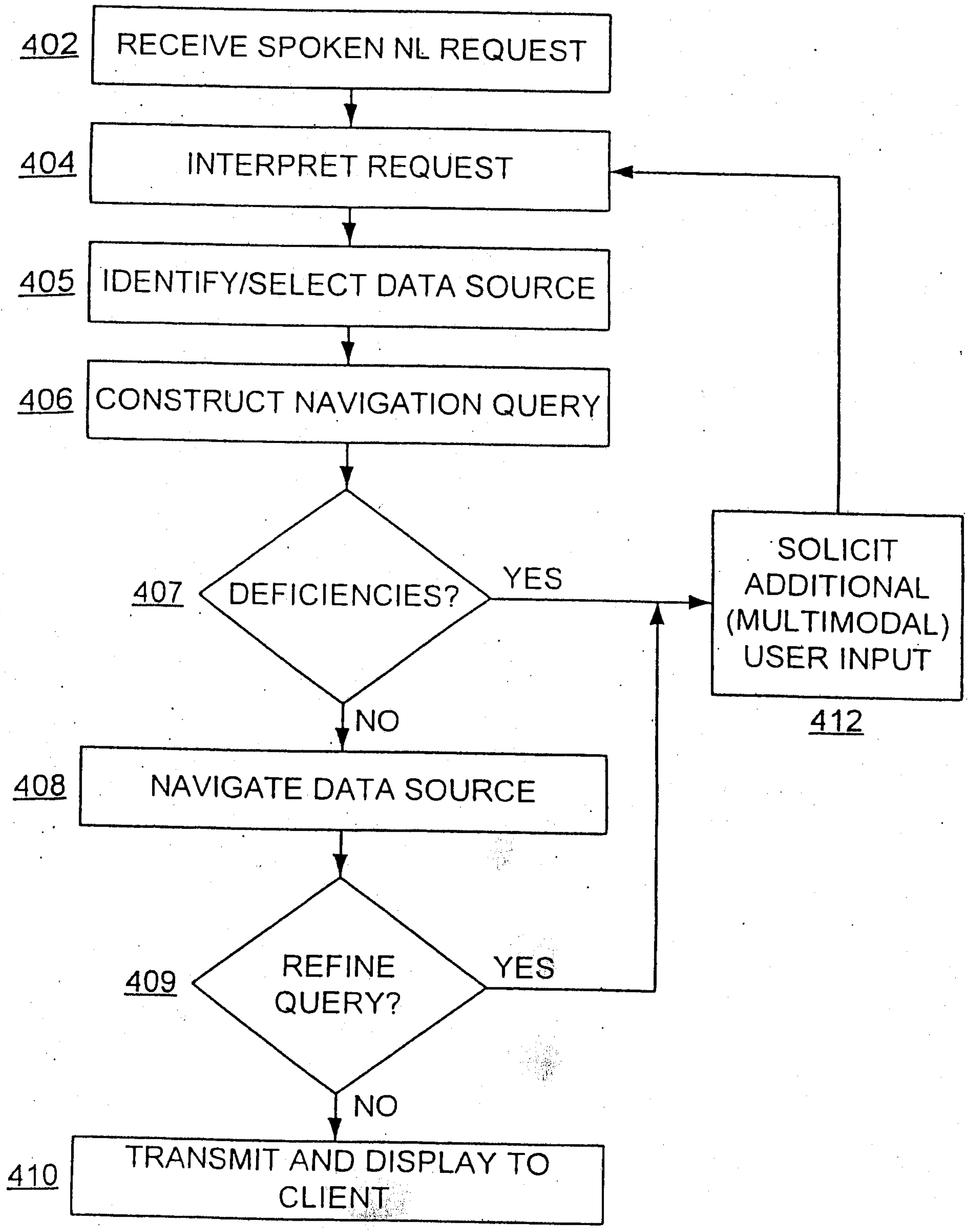

402
RECEIVE SPOKEN NL REQUEST
404
INTERPRET REQUEST
405
IDENTIFY/SELECT DATA SOURCE
406
CONSTRUCT NAVIGATION QUERY
407
DEFICIENCIES?
YES
NO
SOLICIT ADDITIONAL (MULTIMODAL) USER INPUT
412
408
NAVIGATE DATA SOURCE
409
REFINE QUERY?
YES
NO
410
TRANSMIT AND DISPLAY TO CLIENT

Fig. 4

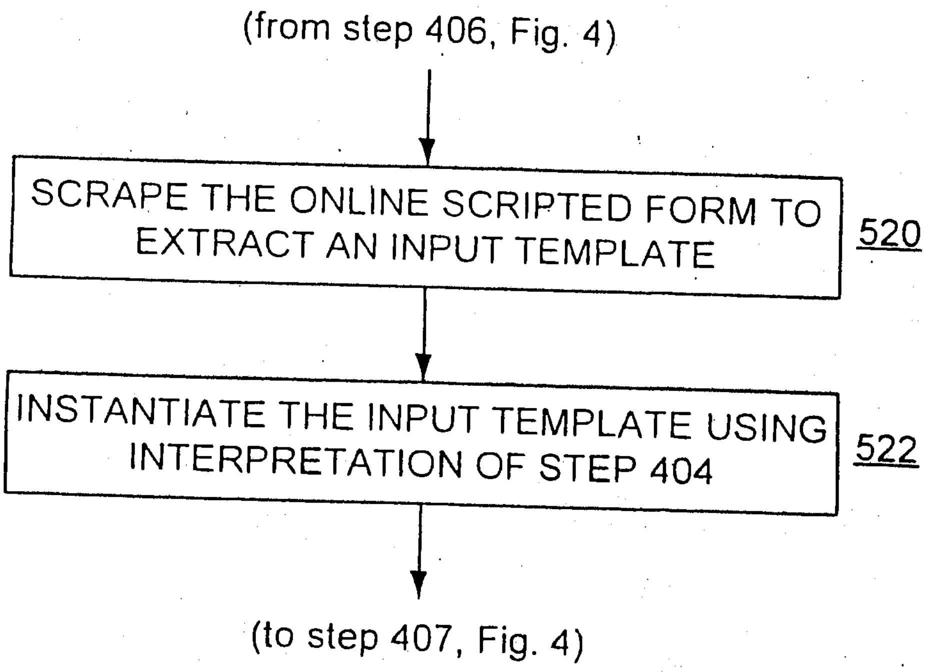

(from step 406, Fig. 4)
SCRAPE THE ONLINE SCRIPTED FORM TO EXTRACT AN INPUT TEMPLATE
520
INSTANTIATE THE INPUT TEMPLATE USING INTERPRETATION OF STEP 404
522
(to step 407, Fig. 4)

Fig. 5

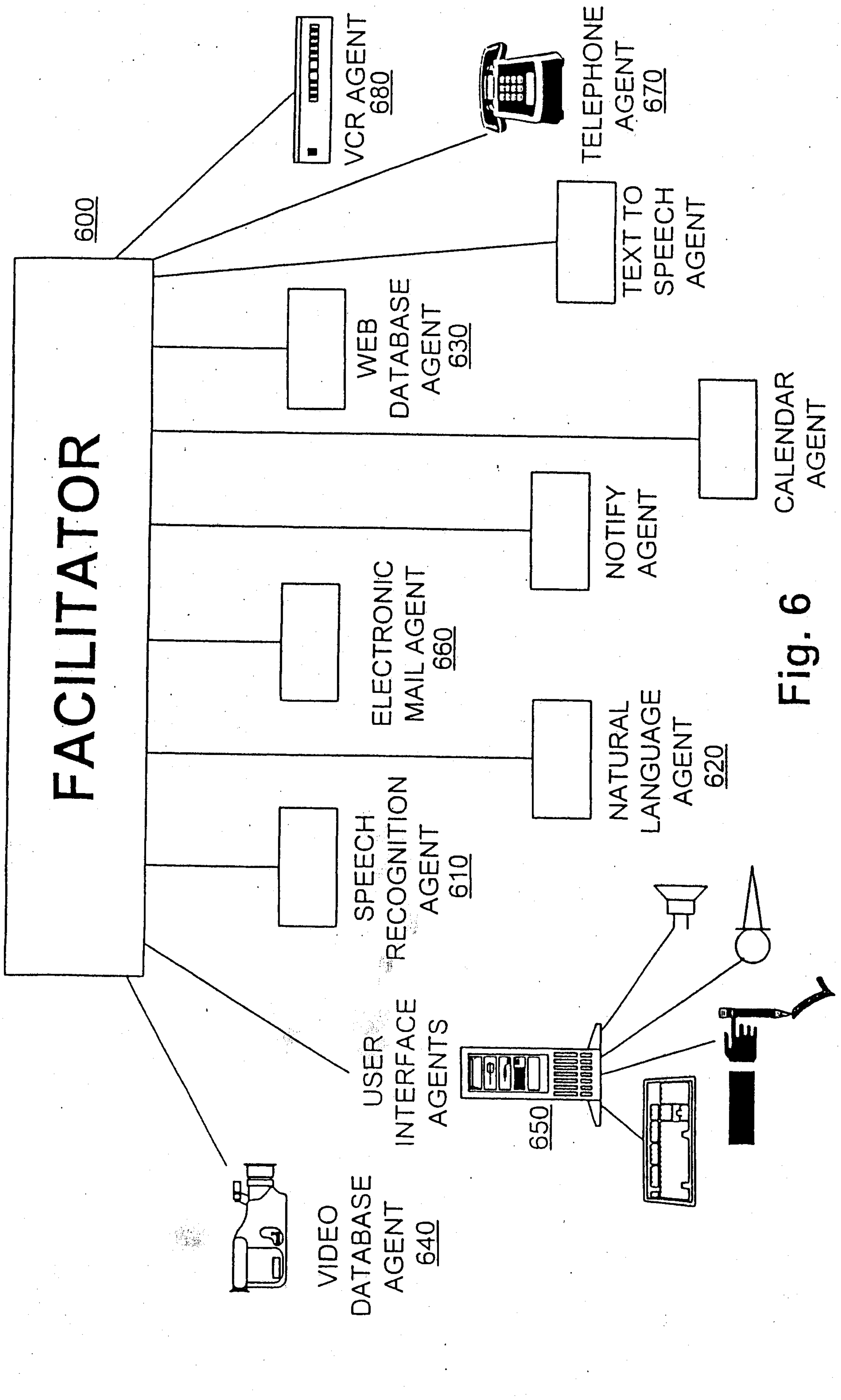

FACILITATOR
600
VCR AGENT
680
TELEPHONE AGENT
670
TEXT TO SPEECH AGENT
WEB DATABASE AGENT
630
CALENDAR AGENT
NOTIFY AGENT
ELECTRONIC MAIL AGENT
660
NATURAL LANGUAGE AGENT
620
SPEECH RECOGNITION AGENT
610
USER INTERFACE AGENTS
650
VIDEO DATABASE AGENT
640

Fig. 6

OIPE

MAY 0 6 2003

# PART B - FEE(S) TRANSMITTAL

**Complete and send this form, together with applicable fee(s), to:** **Mail**
**Box ISSUE FEE**
**Commissioner for Patents**
**Washington, D.C. 20231**
**Fax** **(703)746-4000**

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 4 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as [illegible]ss corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Legibly mark-up with any corrections or use Block 1)

7590 03/11/2003

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY, NJ 07702

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

**Certificate of Mailing or Transmission**

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Box Issue Fee address above, or being facsimile transmitted to the USPTO, on the date indicated below.

Barbara J. Jackson (Depositor's name)
Barbara J Jackson (Signature)
April 30, 2003 (Date)

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 09/608,872 | 06/30/2000 | Christine Halversen | SRILP037B | 2382 |

TITLE OF INVENTION: MOBILE NAVIGATION OF NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT

| APPLN. TYPE | SMALL ENTITY | ISSUE FEE | PUBLICATION FEE | TOTAL FEE(S) DUE | DATE DUE |
|---|---|---|---|---|---|
| nonprovisional | YES | $650 | $0 | $650 | 06/11/2003 |

| EXAMINER | ART UNIT | CLASS-SUBCLASS |
|---|---|---|
| JEAN, FRANTZ B | 2155 | 709-218000 |

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list (1) the names of up to 3 registered patent attorneys or agents OR, alternatively, (2) the name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

Moser, Patterson &
1 Sheridan, LLP
2 Kin-Wah Tong
3 ____________

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. Inclusion of assignee data is only appropriate when an assignment has been previously submitted to the USPTO or is being submitted under separate cover. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE
SRI International

(B) RESIDENCE: (CITY and STATE OR COUNTRY)
Menlo Park, CA

Please check the appropriate assignee category or categories (will not be printed on the patent) ☐ individual ☒ corporation or other private group entity ☐ government

4a. The following fee(s) are enclosed:

☑ Issue Fee
☐ Publication Fee
☒ Advance Order - # of Copies 1

4b. Payment of Fee(s):

☒ A check in the amount of the fee(s) is enclosed.
☐ Payment by credit card. Form PTO-2038 is attached.
☒ The Commissioner is hereby authorized by charge the required fee(s), or credit any overpayment, to Deposit Account Number 20-0782 (enclose an extra copy of this form).

Commissioner for Patents is requested to apply the Issue Fee and Publication Fee (if any) or to re-apply any previously paid issue fee to the application identified above.

(Authorized Signature) (Date) 4/30/03

NOTE: The Issue Fee and Publication Fee (if required) will not be accepted from anyone other than the applicant; a registered attorney or agent; or the assignee or other party in interest as shown by the records of the United States Patent and Trademark Office.

This collection of information is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, Washington, D.C. 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, Washington, DC 20231.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

05/07/2003 SDENBOB2 00000052 09608872

| | | |
|---|---|---|
| 01 FC:2501 | | 650.00 OP |
| 02 FC:8001 | | 3.00 OP |

TRANSMIT THIS FORM WITH FEE(S)

PTOL-85 (REV. 04-02) Approved for use through 01/31/2004. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

AO 120 (Rev. 08/10)

| | |
|---|---|
| TO: **Mail Stop 8**<br>**Director of the U.S. Patent and Trademark Office**<br>**P.O. Box 1450**<br>**Alexandria, VA 22313-1450** | **REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK** |

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the District of Delaware on the following

☐ Trademarks or ☑ Patents. ( ☐ the patent action involves 35 U.S.C. § 292.):

| DOCKET NO. | DATE FILED | U.S. DISTRICT COURT |
|---|---|---|
| | 1/19/2017 | for the District of Delaware |

| PLAINTIFF | DEFENDANT |
|---|---|
| IPA TECHNOLOGIES INC. | SONY CORPORATION, ET AL. |

| PATENT OR TRADEMARK NO. | DATE OF PATENT OR TRADEMARK | HOLDER OF PATENT OR TRADEMARK |
|---|---|---|
| 1 6,742,021 | 5/25/2004 | IPA TECHNOLOGIES INC. |
| 2 6,523,061 | 2/18/2003 | IPA TECHNOLOGIES INC. |
| 3 6,757,718 | 6/29/2004 | IPA TECHNOLOGIES INC. |
| 4 | | |
| 5 | | |

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

| DATE INCLUDED | INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading |
|---|---|
| | |

| PATENT OR TRADEMARK NO. | DATE OF PATENT OR TRADEMARK | HOLDER OF PATENT OR TRADEMARK |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |

In the above—entitled case, the following decision has been rendered or judgement issued:

| DECISION/JUDGEMENT |
|---|
| |

| CLERK | (BY) DEPUTY CLERK | DATE |
|---|---|---|
| | | |

**Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director**
**Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy**

AO 120 (Rev. 08/10)

| | |
|---|---|
| TO: **Mail Stop 8**<br>**Director of the U.S. Patent and Trademark Office**<br>**P.O. Box 1450**<br>**Alexandria, VA 22313-1450** | **REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK** |

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the District of Delaware on the following

☐ Trademarks or ☑ Patents. ( ☐ the patent action involves 35 U.S.C. § 292.):

| DOCKET NO. | DATE FILED | U.S. DISTRICT COURT |
|---|---|---|
| | 3/17/2017 | for the District of Delaware |

| PLAINTIFF | DEFENDANT |
|---|---|
| IPA TECHNOLOGIES INC. | NVIDIA CORPORATION |

| PATENT OR TRADEMARK NO. | DATE OF PATENT OR TRADEMARK | HOLDER OF PATENT OR TRADEMARK |
|---|---|---|
| 1 6,742,021 | 5/25/2004 | IPA TECHNOLOGIES INC. |
| 2 6,523,061 | 2/18/2003 | IPA TECHNOLOGIES INC. |
| 3 6,757,718 | 6/29/2004 | IPA TECHNOLOGIES INC. |
| 4 | | |
| 5 | | |

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

| DATE INCLUDED | INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading | |
|---|---|---|
| **PATENT OR TRADEMARK NO.** | **DATE OF PATENT OR TRADEMARK** | **HOLDER OF PATENT OR TRADEMARK** |
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

| CLERK | (BY) DEPUTY CLERK | DATE |
|---|---|---|
| | | |

**Copy 1—Upon initiation of action, mail this copy to Director Copy 3—Upon termination of action, mail this copy to Director**
**Copy 2—Upon filing document adding patent(s), mail this copy to Director Copy 4—Case file copy**

# PATENT APPLICATION FEE DETERMINATION RECORD

Effective December 29, 1999

Application or Docket Number

## CLAIMS AS FILED - PART I

| | (Column 1) | | (Column 2) |
|---|---|---|---|
| FOR | NUMBER FILED | | NUMBER EXTRA |
| BASIC FEE | | | |
| TOTAL CLAIMS | 27 | minus 20= | * 7 |
| INDEPENDENT CLAIMS | 3 | minus 3 = | * / |
| MULTIPLE DEPENDENT CLAIM PRESENT | | | |

* If the difference in column 1 is less than zero, enter "0" in column 2

SMALL ENTITY TYPE ☐ OR OTHER THAN SMALL ENTITY

| RATE | FEE | | RATE | FEE |
|---|---|---|---|---|
| | 345.00 | OR | | 690.00 |
| X$ 9= | 63 | OR | X$18= | |
| X39= | / | OR | X78= | |
| +130= | / | OR | +260= | |
| TOTAL | 408 | OR | TOTAL | |

## CLAIMS AS AMENDED - PART II

| AMENDMENT A | 1 | (Column 1)<br>CLAIMS REMAINING AFTER AMENDMENT | | (Column 2)<br>HIGHEST NUMBER PREVIOUSLY PAID FOR | (Column 3)<br>PRESENT EXTRA |
|---|---|---|---|---|---|
| | Total | * 27 | Minus | ** 27 | = — |
| | Independent | * 3 | Minus | *** 3 | = — |
| | FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM | | | | |

SMALL ENTITY OR OTHER THAN SMALL ENTITY

| RATE | ADDITIONAL FEE | | RATE | ADDITIONAL FEE |
|---|---|---|---|---|
| X$ 9= | | OR | X$18= | |
| X39= | | OR | X78= | |
| +130= | | OR | +260= | |
| TOTAL ADDIT. FEE | | OR | TOTAL ADDIT. FEE | |

| AMENDMENT B | [illegible] | (Column 1)<br>CLAIMS REMAINING AFTER AMENDMENT | | (Column 2)<br>HIGHEST NUMBER PREVIOUSLY PAID FOR | (Column 3)<br>PRESENT EXTRA |
|---|---|---|---|---|---|
| | Total | * 27 | Minus | ** 27 | = / |
| | Independent | * 3 | Minus | *** 3 | = / |
| | FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM | | | | |

| RATE | ADDITIONAL FEE | | RATE | ADDITIONAL FEE |
|---|---|---|---|---|
| X$ 9= | | OR | X$18= | |
| X39= | | OR | X78= | |
| +130= | | OR | +260= | |
| TOTAL ADDIT. FEE | | OR | TOTAL ADDIT. FEE | |

| AMENDMENT C | | (Column 1)<br>CLAIMS REMAINING AFTER AMENDMENT | | (Column 2)<br>HIGHEST NUMBER PREVIOUSLY PAID FOR | (Column 3)<br>PRESENT EXTRA |
|---|---|---|---|---|---|
| | Total | * | Minus | ** | = |
| | Independent | * | Minus | *** | = |
| | FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM | | | | |

| RATE | ADDITIONAL FEE | | RATE | ADDITIONAL FEE |
|---|---|---|---|---|
| X$ 9= | | OR | X$18= | |
| X39= | | OR | X78= | |
| +130= | | OR | +260= | |
| TOTAL ADDIT. FEE | | OR | TOTAL ADDIT. FEE | |

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20."

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3."

The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

FORM PTO-875
(Rev. [illegible])

Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

*U.S. GPO: 2000-463-413/29044

ISSUE SLIP STAPLE AREA (for additional cross references)

| POSITION | INITIALS | ID NO. | DATE |
|---|---|---|---|
| | | | |
| FEE DETERMINATION | | 62814 | 7/12/00 |
| O.I.P.E. CLASSIFIER | | 48 | 2/27/00 |
| FORMALITY REVIEW | | 60871 | 6/3/00 |
| RESPONSE FORMALITY REVIEW | | | |
| | | 60894 | 11/5/00 |

## INDEX OF CLAIMS

| | | | |
|---|---|---|---|
| ✓ ............................... Rejected | | N ............................... Non-elected | |
| = ............................... Allowed | | I ............................... Interference | |
| — (Through numeral)... Canceled | | A ............................... Appeal | |
| ÷ ............................... Restricted | | O ............................... Objected | |

| Claim | | Date | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Final | Original | | | | | | | | | | |
| | 1 | | | | | | | | | | |
| | 2 | | | | | | | | | | |
| | 3 | | | | | | | | | | |
| | 4 | | | | | | | | | | |
| | 5 | | | | | | | | | | |
| | 6 | | | | | | | | | | |
| | 7 | | | | | | | | | | |
| | 8 | | | | | | | | | | |
| | 9 | | | | | | | | | | |
| | 10 | | | | | | | | | | |
| | 11 | | | | | | | | | | |
| | 12 | | | | | | | | | | |
| | 13 | | | | | | | | | | |
| | 14 | | | | | | | | | | |
| | 15 | | | | | | | | | | |
| | 16 | | | | | | | | | | |
| | 17 | | | | | | | | | | |
| | 18 | | | | | | | | | | |
| | 19 | | | | | | | | | | |
| | 20 | | | | | | | | | | |
| | 21 | | | | | | | | | | |
| | 22 | | | | | | | | | | |
| | 23 | | | | | | | | | | |
| | 24 | | | | | | | | | | |
| | 25 | | | | | | | | | | |
| | 26 | | | | | | | | | | |
| | 27 | | | | | | | | | | |
| | 28 | | | | | | | | | | |
| | 29 | | | | | | | | | | |
| | 30 | | | | | | | | | | |
| | 31 | | | | | | | | | | |
| | 32 | | | | | | | | | | |
| | 33 | | | | | | | | | | |
| | 34 | | | | | | | | | | |
| | 35 | | | | | | | | | | |
| | 36 | | | | | | | | | | |
| | 37 | | | | | | | | | | |
| | 38 | | | | | | | | | | |
| | 39 | | | | | | | | | | |
| | 40 | | | | | | | | | | |
| | 41 | | | | | | | | | | |
| | 42 | | | | | | | | | | |
| | 43 | | | | | | | | | | |
| | 44 | | | | | | | | | | |
| | 45 | | | | | | | | | | |
| | 46 | | | | | | | | | | |
| | 47 | | | | | | | | | | |
| | 48 | | | | | | | | | | |
| | 49 | | | | | | | | | | |
| | 50 | | | | | | | | | | |

| Claim | | Date | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Final | Original | 4/9/01 | 10/2/01 | 2/8/02 | 9/3/02 | 3/7/03 | | | | | |
| | 51 | | | | | — | | | | | |
| | 52 | | | | | — | | | | | |
| | 53 | | | | | — | | | | | |
| | 54 | | | | | — | | | | | |
| | 55 | | | | | — | | | | | |
| 1 | 56 | ✓ | ✓ | ✓ | ✓ | = | | | | | |
| 2 | 57 | | | | | | | | | | |
| 3 | 58 | | | | | | | | | | |
| 4 | 59 | | | | | | | | | | |
| 5 | 60 | | | | | | | | | | |
| 6 | 61 | | | | | | | | | | |
| 7 | 62 | | | | | | | | | | |
| 8 | 63 | | | | | | | | | | |
| 9 | 64 | | | | | | | | | | |
| 10 | 65 | | | | | | | | | | |
| 11 | 66 | | | | | | | | | | |
| 12 | 67 | | | | | | | | | | |
| 13 | 68 | | | | | | | | | | |
| 14 | 69 | | | | | | | | | | |
| 15 | 70 | | | | | | | | | | |
| 16 | 71 | | | | | | | | | | |
| 17 | 72 | | | | | | | | | | |
| 18 | 73 | | | | | | | | | | |
| 19 | 74 | | | | | | | | | | |
| 20 | 75 | | | | | | | | | | |
| 21 | 76 | | | | | | | | | | |
| 22 | 77 | | | | | | | | | | |
| 23 | 78 | | | | | | | | | | |
| 24 | 79 | | | | | | | | | | |
| 25 | 80 | | | | | | | | | | |
| 26 | 81 | | | | | | | | | | |
| 27 | 82 | ✓ | ✓ | ✓ | ✓ | = | | | | | |
| | 83 | | | | | | | | | | |
| | 84 | | | | | | | | | | |
| | 85 | | | | | | | | | | |
| | 86 | | | | | | | | | | |
| | 87 | | | | | | | | | | |
| | 88 | | | | | | | | | | |
| | 89 | | | | | | | | | | |
| | 90 | | | | | | | | | | |
| | 91 | | | | | | | | | | |
| | 92 | | | | | | | | | | |
| | 93 | | | | | | | | | | |
| | 94 | | | | | | | | | | |
| | 95 | | | | | | | | | | |
| | 96 | | | | | | | | | | |
| | 97 | | | | | | | | | | |
| | 98 | | | | | | | | | | |
| | 99 | | | | | | | | | | |
| | 100 | | | | | | | | | | |

| Claim | | Date | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Final | Original | | | | | | | | | | |
| | 101 | | | | | | | | | | |
| | 102 | | | | | | | | | | |
| | 103 | | | | | | | | | | |
| | 104 | | | | | | | | | | |
| | 105 | | | | | | | | | | |
| | 106 | | | | | | | | | | |
| | 107 | | | | | | | | | | |
| | 108 | | | | | | | | | | |
| | 109 | | | | | | | | | | |
| | 110 | | | | | | | | | | |
| | 111 | | | | | | | | | | |
| | 112 | | | | | | | | | | |
| | 113 | | | | | | | | | | |
| | 114 | | | | | | | | | | |
| | 115 | | | | | | | | | | |
| | 116 | | | | | | | | | | |
| | 117 | | | | | | | | | | |
| | 118 | | | | | | | | | | |
| | 119 | | | | | | | | | | |
| | 120 | | | | | | | | | | |
| | 121 | | | | | | | | | | |
| | 122 | | | | | | | | | | |
| | 123 | | | | | | | | | | |
| | 124 | | | | | | | | | | |
| | 125 | | | | | | | | | | |
| | 126 | | | | | | | | | | |
| | 127 | | | | | | | | | | |
| | 128 | | | | | | | | | | |
| | 129 | | | | | | | | | | |
| | 130 | | | | | | | | | | |
| | 131 | | | | | | | | | | |
| | 132 | | | | | | | | | | |
| | 133 | | | | | | | | | | |
| | 134 | | | | | | | | | | |
| | 135 | | | | | | | | | | |
| | 136 | | | | | | | | | | |
| | 137 | | | | | | | | | | |
| | 138 | | | | | | | | | | |
| | 139 | | | | | | | | | | |
| | 140 | | | | | | | | | | |
| | 141 | | | | | | | | | | |
| | 142 | | | | | | | | | | |
| | 143 | | | | | | | | | | |
| | 144 | | | | | | | | | | |
| | 145 | | | | | | | | | | |
| | 146 | | | | | | | | | | |
| | 147 | | | | | | | | | | |
| | 148 | | | | | | | | | | |
| | 149 | | | | | | | | | | |
| | 150 | | | | | | | | | | |

If more than 150 claims or 10 actions
staple additional sheet here

(LEFT INSIDE)

# SEARCHED

| Class | Sub. | Date | Exmr. |
|---|---|---|---|
| 709 | 202 | 4/10/01 | F.B |
| 709 | 218 | | |
| 707 | 5<br>3<br>4 | | |
| 704 | 257 | | |
| updated | | 9/30/02 | |
| updated | | 3/7/03 | |
| 704 | 270.1<br>275<br>246<br>257 | 3/7/03 | |
| 709 | 217<br>219<br>227 | | |

# SEARCH NOTES
## (INCLUDING SEARCH STRATEGY)

| | Date | Exmr. |
|---|---|---|
| West SEARCH | 4/10/01 | F.B |
| West, DerWent<br>TDBD, EPO, JPO | 9/29/02<br>9/30/02 | |
| consulted with David Wiley | 9/30/02 | |
| updated | 3/7/03 | |
| West, DerWent TDBD, EPO, JPO, PG Pub<br>inventor search<br>check for possible Double patenting | 3/7/03 | |
| NPL | 3/7/03 | |

# INTERFERENCE SEARCHED

| Class | Sub. | Date | Exmr. |
|---|---|---|---|
| 709 | 202<br>217<br>218<br>219<br>227 | 3/7/03 | |
| 704 | 257 | | |

US006742021B1

(12) **United States Patent**
**Halverson et al.**

(10) **Patent No.: US 6,742,021 B1**
(45) **Date of Patent: May 25, 2004**

---

(54) **NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT WITH MULTIMODAL ERROR FEEDBACK**

(75) Inventors: **Christine Halverson**, San Jose, CA (US); **Luc Julia**, Menlo Park, CA (US); **Dimitris Voutsas**, Thessaloniki (GR); **Aden J. Cheyer**, Palo Alto, CA (US)

(73) Assignee: **SRI International, Inc.**, Menlo Park, CA (US)

( * ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/524,095**

(22) Filed: **Mar. 13, 2000**

**Related U.S. Application Data**

(63) Continuation-in-part of application No. 09/225,198, filed on Jan. 5, 1999.

(60) Provisional application No. 60/124,718, filed on Mar. 17, 1999, provisional application No. 60/124,720, filed on Mar. 17, 1999, and provisional application No. 60/124,719, filed on Mar. 17, 1999.

(51) **Int. Cl.**$^7$ ................................................ **G06F 15/16**

(52) **U.S. Cl.** ............................... **709/218**; 707/5; 707/4; 707/102

(58) **Field of Search** .............................. 709/218; 707/5, 707/4, 102; 704/257, 231

(56) **References Cited**

U.S. PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| 5,197,005 A | 3/1993 | Schwartz et al. ........... | 364/419 |
| 5,386,556 A | 1/1995 | Hedin et al. ................ | 395/600 |

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

| | | |
|---|---|---|
| EP | 0 803 826 A2 | 10/1997 |
| WO | WO 00/11869 | 3/2000 |

OTHER PUBLICATIONS

http://www.ai.sri.com/~lesaf/commandtalk.html: "CommandTalk: A Spoken–Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, DC, pp. 1–7, Association for Computational Linguistics.

"The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty–Seventh Annual Meeting of the ACL", pp. 183–190, University of Maryland, College Park, MD, Association for Computational Linguistics.

Stent, Amanda et al., "The CommandTalk Spoken Dialogue System", SRI International.

Moore, Robert et al., "CommandTalk: A Spoken–Language Interface for Battlefield Simulations", Oct. 23, 1997, SRI International.

Dowding, John et al., "Interpreting Language in Context in CommandTalk", Feb. 5, 1999, SRI International.

http://www.ai.sri.com/~oaa/infowiz.html, InfoWiz: An Animated Voice Interactive Information System, May 8, 2000.

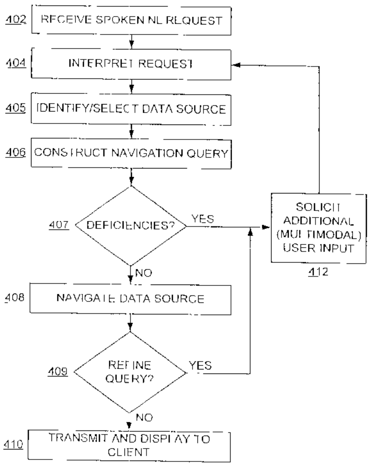
(List continued on next page.)

*Primary Examiner*—James P. Trammell
*Assistant Examiner*—Firmin Backer
(74) *Attorney, Agent, or Firm*—Moser, Patterson & Sheridan, LLP.; Kin-Wah Tong, Esq.

(57) **ABSTRACT**

A system, method, and article of manufacture are provided for navigating an electronic data source by means of spoken language. When a spoken input request is received from a user, it is interpreted. Additional input is solicited from the user in a modality different than the original request and used to refine the navigation query. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources.

**132 Claims, 7 Drawing Sheets**



U.S. PATENT DOCUMENTS

| | | | | | |
|---|---|---|---|---|---|
| 5,434,777 | A | | 7/1995 | Luciw | 364/419 |
| 5,519,608 | A | | 5/1996 | Kupiec | 364/419.08 |
| 5,608,624 | A | | 3/1997 | Luciw | 395/794 |
| 5,721,938 | A | | 2/1998 | Stuckey | 395/754 |
| 5,729,659 | A | | 3/1998 | Potter | 395/2.79 |
| 5,748,974 | A | | 5/1998 | Johnson | 395/759 |
| 5,774,859 | A | | 6/1998 | Houser et al. | 704/275 |
| 5,794,050 | A | | 8/1998 | Dahlgren et al. | 395/708 |
| 5,802,526 | A | | 9/1998 | Fawcett et al. | 707/104 |
| 5,805,775 | A | | 9/1998 | Eberman et al. | 395/12 |
| 5,855,002 | A | | 12/1998 | Armstrong | 704/270 |
| 5,890,123 | A | | 3/1999 | Brown et al. | 704/275 |
| 5,963,940 | A | | 10/1999 | Liddy et al. | 707/5 |
| 6,003,072 | A | | 12/1999 | Gerritsen et al. | 709/218 |
| 6,012,030 | A | | 1/2000 | French-St. George et al. | 704/275 |
| 6,021,427 | A | | 1/2000 | Spagna et al. | |
| 6,026,388 | A | | 2/2000 | Liddy et al. | 707/1 |
| 6,080,202 | A | | 6/2000 | Strickland et al. | |
| 6,144,989 | A | | 11/2000 | Hodjat et al. | |
| 6,173,279 | B1 | * | 1/2001 | Levin et al. | 707/5 |
| 6,192,338 | B1 | * | 2/2001 | Zasto et al. | 704/257 |
| 6,226,666 | B1 | | 5/2001 | Chang et al. | |
| 6,338,081 | B1 | | 1/2002 | Furusawa et al. | |

OTHER PUBLICATIONS

Dowding, John, "Interleaving Syntax and Semantics in an Efficient Bottom–up Parser", SRI International.

Moore, Robert et al., "Combining Linguistic and Statistical Knowledge Sources in Natural–Language Processing for ATIS", SRI International.

Dowding, John et al., "Gemini: A Natural Language System For Spoken–Language Understanding", SRI International.

Moran, Douglas B. et al., "Intelligent Agent–based User Interfaces", Article Intelligence center, SRI International.

Martin, David L. et al., "Building Distributed Software Systems with the Open Agent Architecture".

Julia, Luc. et al., "Cooperative Agents and Recognition System (CARS) for Drivers and Passengers"; SRI International.

Moran, Douglas et al., "Multimodal User Interfaces in the Open Agent Architecture".

Cheyer, Adam et al., "Multimodal Maps: An Agent–based Approach", SRI International.

Cutkosky, Mark R. et al., "An Experiment in Integrating Concurrent Engineering Systems".

Martin, David et al., "Development Tools for the Open Agent Architecture", The Practical Application of Intelleigent Agents and Multi–Agent Technology (PAAM96), London, Apr. 1996.

Cheyer, Adam et al., "The Open Agent Architecture$_{tm}$", SRI International, AI center.

Dejima, Inc., http://www.dejima.com/.

Cohen, Philip et al., "An Open Agent Architecture", AAAI Spring Symposium, pp. 1–8, Mar. 1994.

Martin, David et al., "Information Brokering in an Agent Architecture", Proceeding of the $2^{nd}$ Int'l Conference on Practical Application of Intelligent Agents & Multi–Agent Technology, London, Apr. 1997.

* cited by examiner

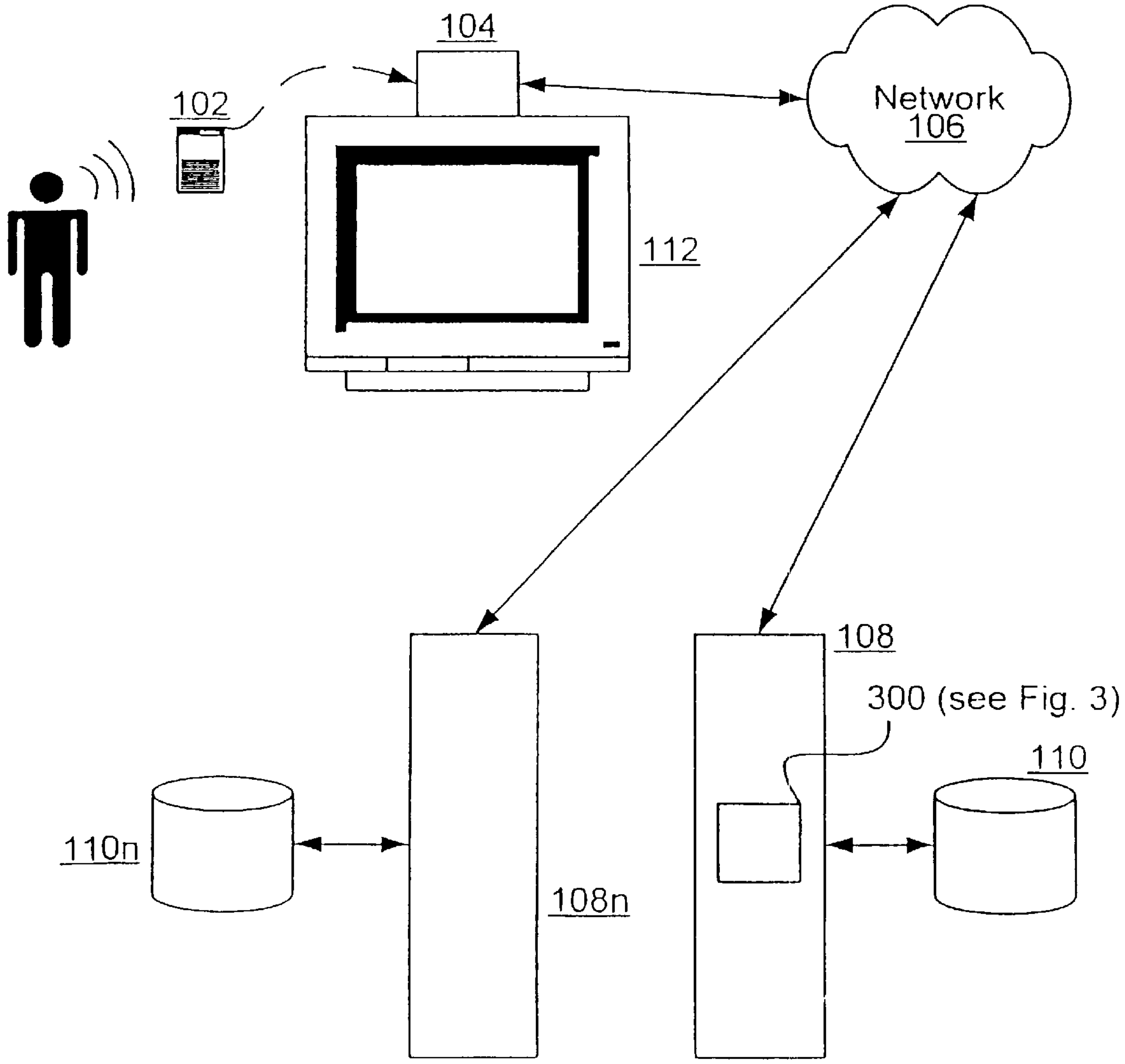


Fig. 1a

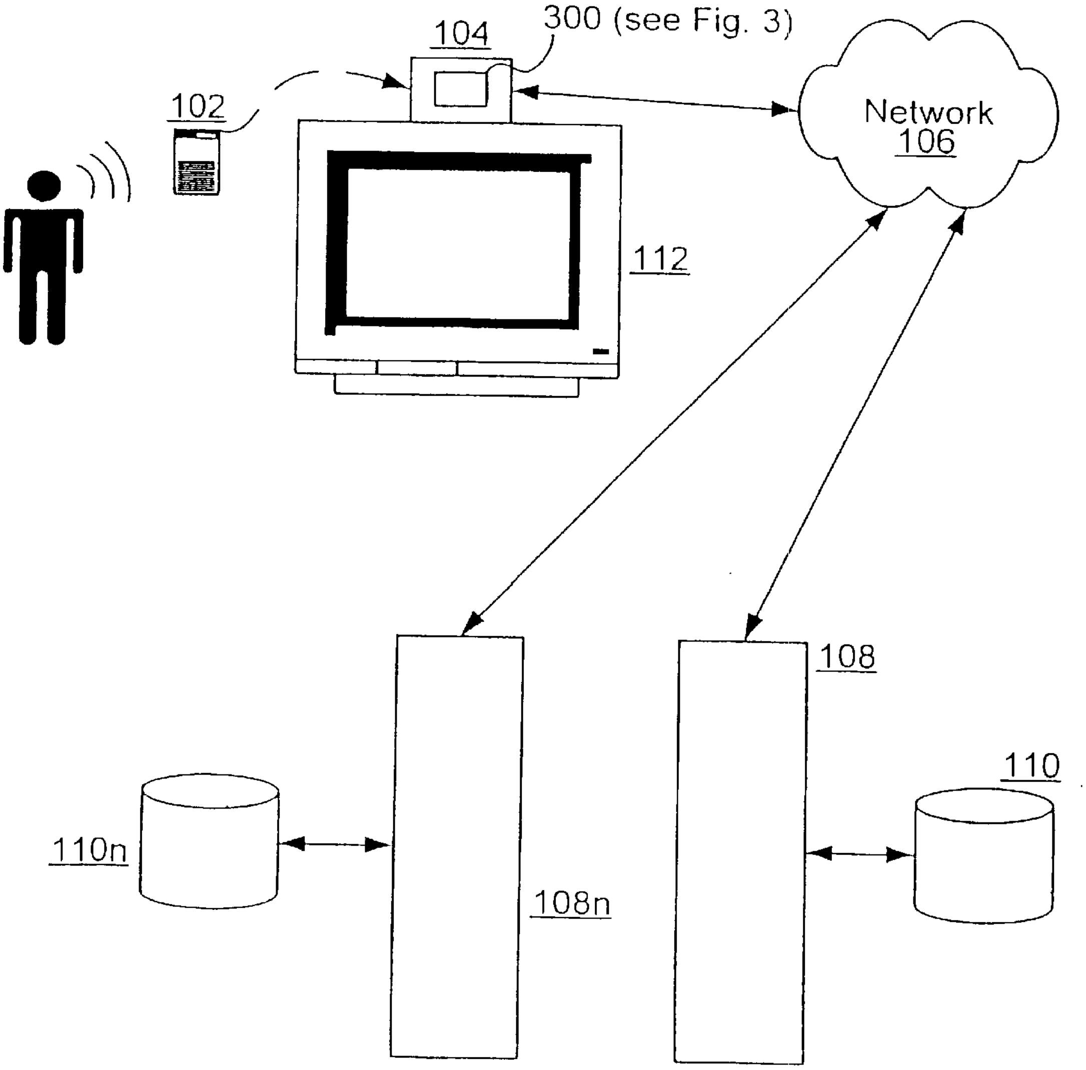


Fig. 1b

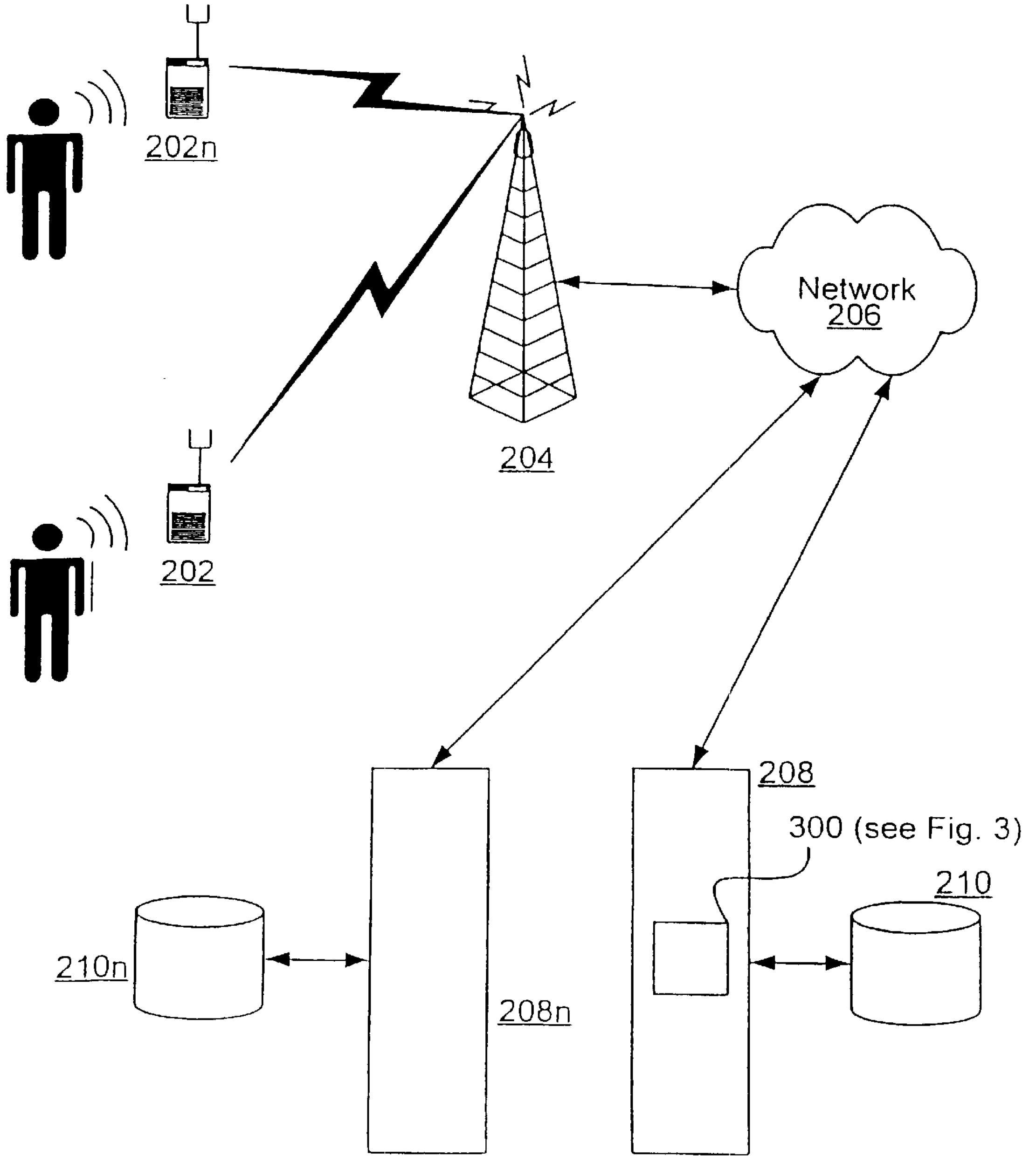


Fig. 2

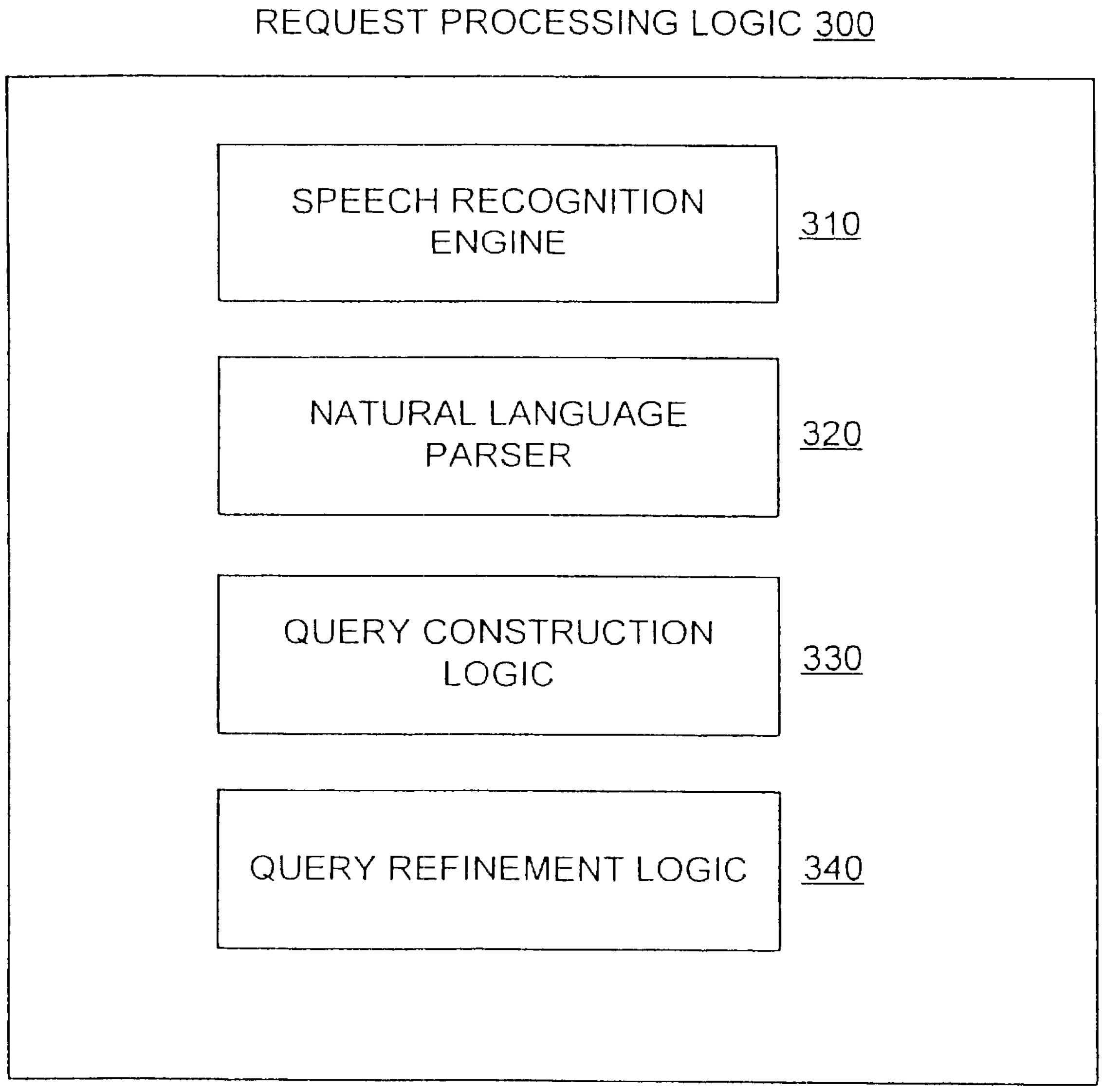

REQUEST PROCESSING LOGIC 300
SPEECH RECOGNITION ENGINE
310
NATURAL LANGUAGE PARSER
320
QUERY CONSTRUCTION LOGIC
330
QUERY REFINEMENT LOGIC
340

Fig. 3

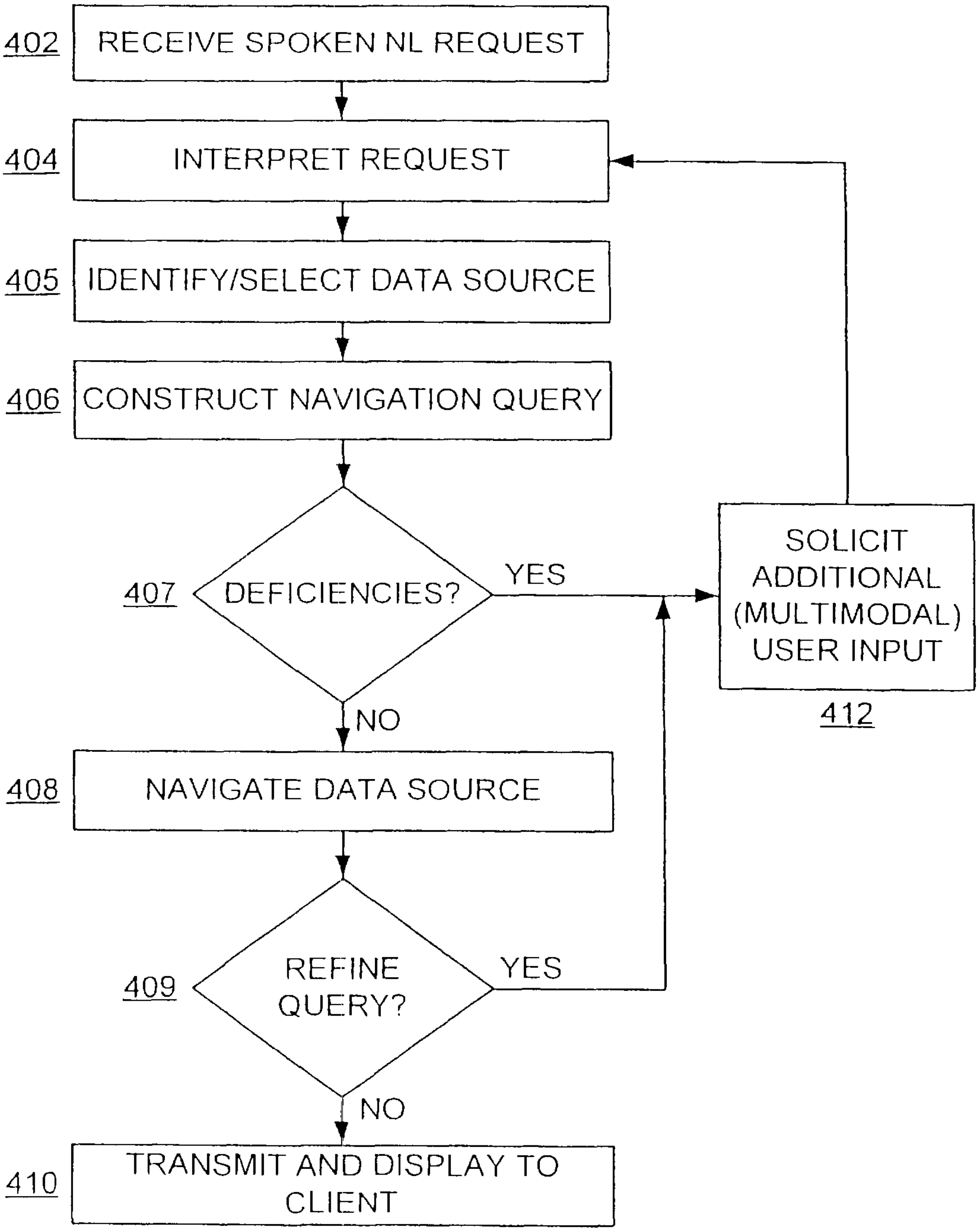

402
RECEIVE SPOKEN NL REQUEST
404
INTERPRET REQUEST
405
IDENTIFY/SELECT DATA SOURCE
406
CONSTRUCT NAVIGATION QUERY
407
DEFICIENCIES?
YES
NO
SOLICIT ADDITIONAL (MULTIMODAL) USER INPUT
412
408
NAVIGATE DATA SOURCE
409
REFINE QUERY?
YES
NO
410
TRANSMIT AND DISPLAY TO CLIENT

Fig. 4

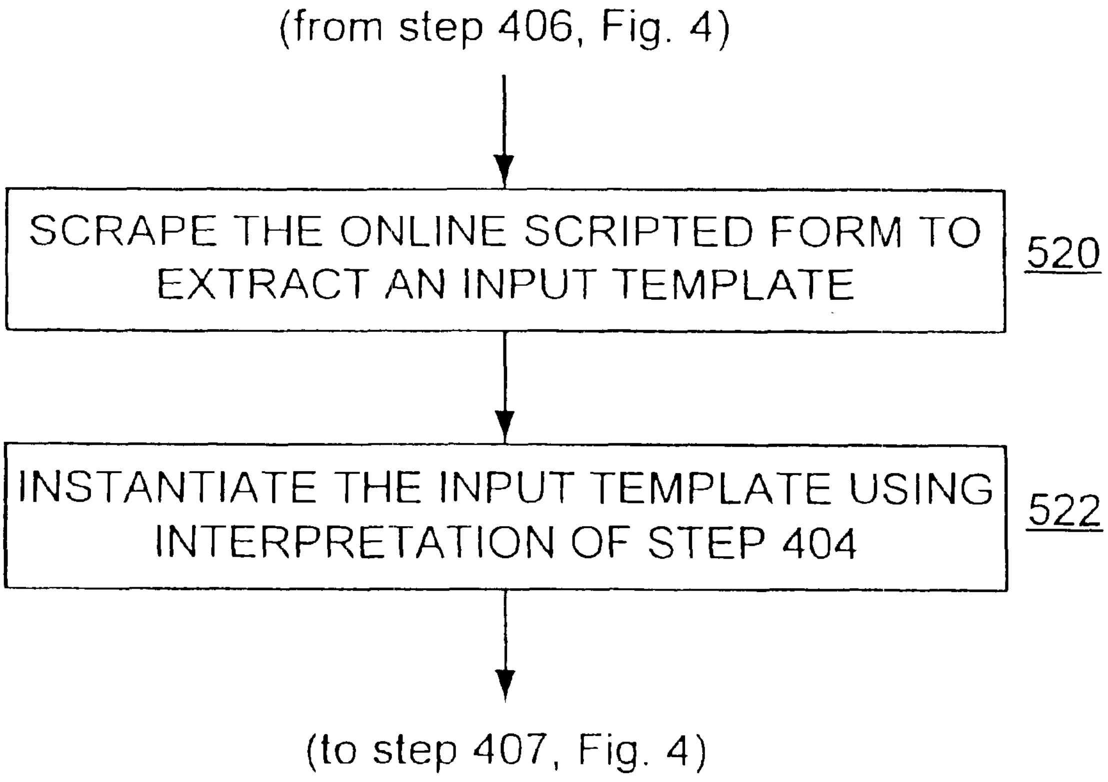

(from step 406, Fig. 4)
SCRAPE THE ONLINE SCRIPTED FORM TO EXTRACT AN INPUT TEMPLATE
520
INSTANTIATE THE INPUT TEMPLATE USING INTERPRETATION OF STEP 404
522
(to step 407, Fig. 4)

Fig. 5

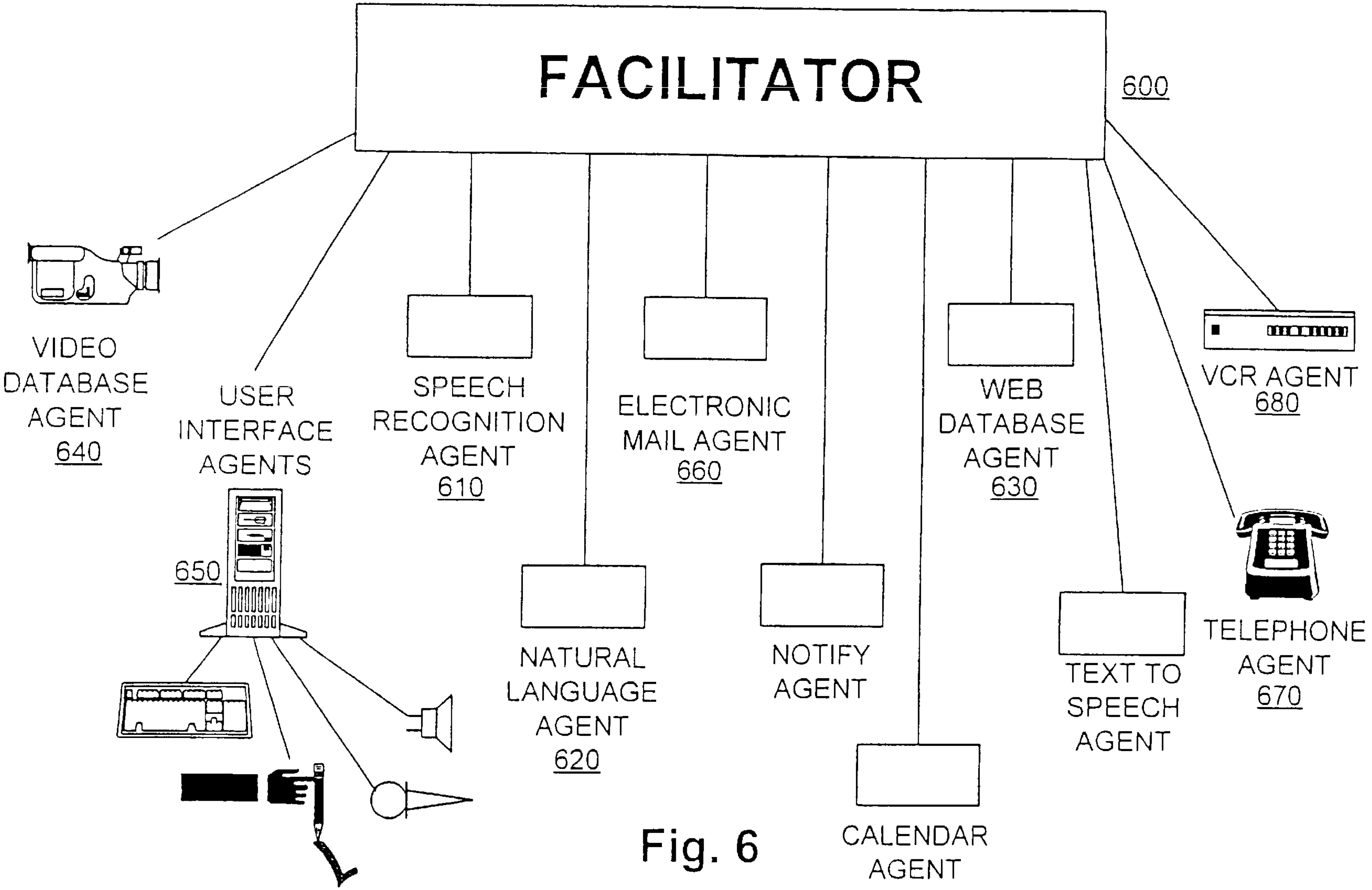

FACILITATOR
600
VIDEO DATABASE AGENT 640
USER INTERFACE AGENTS
650
SPEECH RECOGNITION AGENT 610
NATURAL LANGUAGE AGENT 620
ELECTRONIC MAIL AGENT 660
NOTIFY AGENT
CALENDAR AGENT
WEB DATABASE AGENT 630
TEXT TO SPEECH AGENT
VCR AGENT 680
TELEPHONE AGENT 670

Fig. 6

# NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT WITH MULTIMODAL ERROR FEEDBACK

This is a Continuation In Part of co-pending U.S. patent application Ser. No. 09/225,198, filed Jan. 5, 1999, Provisional U.S. patent application Ser. No. 60/124,718, filed Mar. 17, 1999, Provisional U.S. patent application Ser. No. 60/124,720, filed Mar. 17, 1999, and Provisional U.S. patent application Ser. No. 60/124,719, filed Mar. 17, 1999, from which applications priority is claimed and these application are incorporated herein by reference.

## BACKGROUND OF THE INVENTION

The present invention relates generally to the navigation of electronic data by means of spoken natural language requests, and to feedback mechanisms and methods for resolving the errors and ambiguities that may be associated with such requests.

As global electronic connectivity continues to grow, and the universe of electronic data potentially available to users continues to expand, there is a growing need for information navigation technology that allows relatively naïve users to navigate and access desired data by means of natural language input. In many of the most important markets—including the home entertainment arena, as well as mobile computing—spoken natural language input is highly desirable, if not ideal. As just one example, the proliferation of high-bandwidth communications infrastructure for the home entertainment market (cable, satellite, broadband) enables delivery of movies-on-demand and other interactive multimedia content to the consumer's home television set. For users to take full advantage of this content stream ultimately requires interactive navigation of content databases in a manner that is too complex for user-friendly selection by means of a traditional remote-control clicker. Allowing spoken natural language requests as the input modality for rapidly searching and accessing desired content is an important objective for a successful consumer entertainment product in a context offering a dizzying range of database content choices. As further examples, this same need to drive navigation of (and transaction with) relatively complex data warehouses using spoken natural language requests applies equally to surfing the Internet/Web or other networks for general information, multimedia content, or e-commerce transactions.

In general, the existing navigational systems for browsing electronic databases and data warehouses (search engines, menus, etc.), have been designed without navigation via spoken natural language as a specific goal. So today's world is full of existing electronic data navigation systems that do not assume browsing via natural spoken commands, but rather assume text and mouse-click inputs (or in the case of TV remote controls, even less). Simply recognizing voice commands within an extremely limited vocabulary and grammar—the spoken equivalent of button/click input (e.g., speaking "channel 5" selects TV channel 5)—is really not sufficient by itself to satisfy the objectives described above. In order to deliver a true "win" for users, the voice-driven front-end must accept spoken natural language input in a manner that is intuitive to users. For example, the front-end should not require learning a highly specialized command language or format. More fundamentally, the front-end must allow users to speak directly in terms of what the user ultimately wants —e.g., "I'd like to see a Western film directed by Clint Eastwood"—as opposed to speaking in terms of arbitrary navigation structures (e.g., hierarchical layers of menus, commands, etc.) that are essentially artifacts reflecting constraints of the pre-existing text/click navigation system. At the same time, the front-end must recognize and accommodate the reality that a stream of naive spoken natural language input will, over time, typically present a variety of errors and/or ambiguities: e.g., garbled/unrecognized words (did the user say "Eastwood" or "Easter"?) and under-constrained requests ("Show me the Clint Eastwood movie"). An approach is needed for handling and resolving such errors and ambiguities in a rapid, user-friendly, non-frustrating manner.

What is needed is a methodology and apparatus for rapidly constructing a voice-driven front-end atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the step-by-step browsing architecture of the existing navigation system, and wherein any errors or ambiguities in user input are rapidly and conveniently resolved. The solution to this need should be compatible with the constraints of a multi-user, distributed environment such as the Internet/Web or a proprietary high-bandwidth content delivery network; a solution contemplating one-at-a-time user interactions at a single location is insufficient, for example.

## SUMMARY OF THE INVENTION

The present invention addresses the above needs by providing a system, method, and article of manufacture for navigating network-based electronic data sources in response to spoken input requests. When a spoken input request is received from a user, it is interpreted, such as by using a speech recognition engine to extract speech data from acoustic voice signals, and using a language parser to linguistically parse the speech data. The interpretation of the spoken request can be performed on a computing device locally with the user or remotely from the user. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is then transmitted to a client device of the user. If the network data source is a database, the navigation query is constructed in the format of a database query language.

Typically, errors or ambiguities emerge in the interpretation of the spoken request, such that the system cannot instantiate a complete, valid navigational template. This is to be expected occasionally, and one preferred aspect of the invention is the ability to handle such errors and ambiguities in relatively graceful and user-friendly manner. Instead of simply rejecting such input and defaulting to traditional input modes or simply asking the user to try again, a preferred embodiment of the present invention seeks to converge rapidly toward instantiation of a valid navigational template by soliciting additional clarification from the user as necessary, either before or after a navigation of the data source, via multimodal input, i.e., by means of menu selection or other input modalities including and in addition to spoken input. This clarifying, multi-modal dialogue takes advantage of whatever partial navigational information has been gleaned from the initial interpretation of the user's spoken request. This clarification process continues until the system converges toward an adequately instantiated navigational template, which is in turn used to navigate the network-based data and retrieve the user's desired information. The retrieved information is transmitted across the network and presented to the user on a suitable client display device.

In a further aspect of the present invention, the construction of the navigation query includes extracting an input template for an online scripted interface to the data source and using the input template to construct the navigation query. The extraction of the input template can include dynamically scraping the online scripted interface.

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention, together with further advantages thereof, may best be understood by reference to the following description taken in conjunction with the accompanying drawings in which:

FIG. **1***a* illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with server-side processing of requests;

FIG. **1***b* illustrates another system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with client-side processing of requests;

FIG. **2** illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention for a mobile computing scenario;

FIG. **3** illustrates the functional logic components of a request processing module in accordance with an embodiment of the present invention;

FIG. **4** illustrates a process utilizing spoken natural language for navigating an electronic database in accordance with one embodiment of the present invention;

FIG. **5** illustrates a process for constructing a navigational query for accessing an online data source via an interactive, scripted (e.g., CGI) form; and

FIG. **6** illustrates an embodiment of the present invention utilizing a community of distributed, collaborating electronic agents.

## DETAILED DESCRIPTION OF THE INVENTION

### 1. System Architecture

### a. Server-End Processing of Spoken Input

FIG. **1***a* is an illustration of a data navigation system driven by spoken natural language input, in accordance with one embodiment of the present invention. As shown, a user's voice input data is captured by a voice input device **102**, such as a microphone. Preferably voice input device **102** includes a button or the like that can be pressed or held-down to activate a listening mode, so that the system need not continually pay attention to, or be confused by, irrelevant background noise. In one preferred embodiment well-suited for the home entertainment setting, voice input device **102** is a portable remote control device with an integrated microphone, and the voice data is transmitted from device **102** preferably via infrared (or other wireless) link to communications box **104** (e.g., a set-top box or a similar communications device that is capable of retransmitting the raw voice data and/or processing the voice data) local to the user's environment and coupled to communications network **106**. The voice data is then transmitted across network **106** to a remote server or servers **108**. The voice data may preferably be transmitted in compressed digitized form, or alternatively—particularly where bandwidth constraints are significant—in analog format (e.g., via frequency modulated transmission), in the latter case being digitized upon arrival at remote server **108**.

At remote server **108**, the voice data is processed by request processing logic **300** in order to understand the user's request and construct an appropriate query or request for navigation of remote data source **110**, in accordance with the interpretation process exemplified in FIG. **4** and FIG. **5** and discussed in greater detail below. For purposes of executing this process, request processing logic **300** comprises functional modules including speech recognition engine **310**, natural language (NL) parser **320**, query construction logic **330**, and query refinement logic **340**, as shown in FIG. **3**. Data source **110** may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably resides on a central server or servers—which may or may not be the same as server **108**, depending on the storage and bandwidth needs of the application and the resources available to the practitioner. Data source **110** may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source **110** are navigated—i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user—using the processes of FIGS. **4** and **5** as described in greater detail below.

Once the desired information has been retrieved from data source **110**, it is electronically transmitted via network **106** to the user for viewing on client display device **112**. In a preferred embodiment well-suited for the home entertainment setting, display device **112** is a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device **112** is coupled to or integrated with a communications box (which is preferably the same as communications box **104**, but may also be a separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network **106**.

Network **106** is a two-way electronic communications network and may be embodied in electronic communication infrastructure including coaxial (cable television) lines, DSL, fiber-optic cable, traditional copper wire (twisted pair), or any other type of hardwired connection. Network **106** may also include a wireless connection such as a satellite-based connection, cellular connection, or other type of wireless connection. Network **106** may be part of the Internet and may support TCP/IP communications, or may be embodied in a proprietary network, or in any other electronic communications network infrastructure, whether packet-switched or connection-oriented. A design consideration is that network **106** preferably provide suitable bandwidth depending upon the nature of the content anticipated for the desired application.

### b. Client-End Processing of Spoken Input

FIG. **1***b* is an illustration of a data navigation system driven by spoken natural language input, in accordance with a second embodiment of the present invention. Again, a user's voice input data is captured by a voice input device **102**, such as a microphone. In the embodiment shown in FIG. **1***b*, the voice data is transmitted from device **202** to requests processing logic **300**, hosted on a local speech processor, for processing and interpretation. In the preferred embodiment illustrated in FIG. **1***b*, the local speech processor is conveniently integrated as part of communications box **104**, although implementation in a physically separate (but communicatively coupled) unit is also possible as will be readily apparent to those of skill in the art. The voice data is processed by the components of request processing logic

**300** in order to understand the user's request and construct an appropriate query or request for navigation of remote data source **110**, in accordance with the interpretation process exemplified in FIGS. **4** and **5** as discussed in greater detail below.

The resulting navigational query is then transmitted electronically across network **106** to data source **110**, which preferably resides on a central server or servers **108**. As in FIG. **1***a*, data source **110** may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source **110** are then navigated—i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user—preferably using the process of FIGS. **4** and **5** as described in greater detail below. Once the desired information has been retrieved from data source **110**, it is electronically transmitted via network **106** to the user for viewing on client display device **112**.

In one embodiment in accordance with FIG. **1***b* and well-suited for the home entertainment setting, voice input device **102** is a portable remote control device with an integrated microphone, and the voice data is transmitted from device **102** preferably via infrared (or other wireless) link to the local speech processor. The local speech processor is coupled to communications network **106**, and also preferably to client display device **112** (especially for purposes of query refinement transmissions, as discussed below in connection with FIG. **4**, step **412**), and preferably may be integrated within or coupled to communications box **104**. In addition, especially for purposes of a home entertainment application, display device **112** is preferably a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device **112** is coupled to a communications box (which is preferably the same as communications box **104**, but may also be a physically separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network **106**.

Design considerations favoring server-side processing and interpretation of spoken input requests, as exemplified in FIG. **1***a*, include minimizing the need to distribute costly computational hardware and software to all client users in order to perform speech and language processing. Design considerations favoring client-side processing, as exemplified in FIG. **1***b*, include minimizing the quantity of data sent upstream across the network from each client, as the speech recognition is performed before transmission across the network and only the query data and/or request needs to be sent, thus reducing the upstream bandwidth requirements.

### c. Mobile Client Embodiment

A mobile computing embodiment of the present invention may be implemented by practitioners as a variation on the embodiments of either FIG. **1***a* or FIG. **1***b*. For example, as depicted in FIG. **2**, a mobile variation in accordance with the server-side processing architecture illustrated in FIG. **1** a may be implemented by replacing voice input device **102**, communications box **104**, and client display device **112**, with an integrated, mobile, information appliance **202** such as a cellular telephone or wireless personal digital assistant (wireless PDA). Mobile information appliance **202** essentially performs the functions of the replaced components. Thus, mobile information appliance **202** receives spoken natural language input requests from the user in the form of voice data, and transmits that data (preferably via wireless data receiving station **204**) across communications network **206** for server-side interpretation of the request, in similar fashion as described above in connection with FIG. **1**. Navigation of data source **210** and retrieval of desired information likewise proceeds in an analogous manner as described above. Display information transmitted electronically back to the user across network **206** is displayed for the user on the display of information appliance **202**, and audio information is output through the appliance's speakers.

Practitioners will further appreciate, in light of the above teachings, that if mobile information appliance **202** is equipped with sufficient computational processing power, then a mobile variation of the client-side architecture exemplified in FIG. **2** may similarly be implemented. In that case, the modules corresponding to request processing logic **300** would be embodied locally in the computational resources of mobile information appliance **202**, and the logical flow of data would otherwise follow in a manner analogous to that previously described in connection with FIG. **1***b*.

As illustrated in FIG. **2**, multiple users, each having their own client input device, may issue requests, simultaneously or otherwise, for navigation of data source **210**. This is equally true (though not explicitly drawn) for the embodiments depicted in FIGS. **1***a* and **1***b*. Data source **210** (or **100**), being a network accessible information resource, has typically already been constructed to support access requests from simultaneous multiple network users, as known by practitioners of ordinary skill in the art. In the case of server-side speech processing, as exemplified in FIGS. **1***a* and **2**, the interpretation logic and error correction logic modules are also preferably designed and implemented to support queuing and multi-tasking of requests from multiple simultaneous network users, as will be appreciated by those of skill in the art.

It will be apparent to those skilled in the art that additional implementations, permutations and combinations of the embodiments set forth in FIGS. **1***a*, **1***b*, and **2** may be created without straying from the scope and spirit of the present invention. For example, practitioners will understand, in light of the above teachings and design considerations, that it is possible to divide and allocate the functional components of request processing logic **300** between client and server. For example, speech recognition—in entirety, or perhaps just early stages such as feature extraction—might be performed locally on the client end, perhaps to reduce bandwidth requirements, while natural language parsing and other necessary processing might be performed upstream on the server end, so that more extensive computational power need not be distributed locally to each client. In that case, corresponding portions of request processing logic **300**, such as speech recognition engine **310** or portions thereof, would reside locally at the client as in FIG. **1***b*, while other component modules would be hosted at the server end as in FIGS. **1***a* and **2**.

Further, practitioners may choose to implement the each of the various embodiments described above on any number of different hardware and software computing platforms and environments and various combinations thereof, including, by way of just a few examples: a general-purpose hardware microprocessor such as the Intel Pentium series; operating system software such as Microsoft Windows/CE, Palm OS, or Apple Mac OS (particularly for client devices and client-side processing), or Unix, Linux, or Windows/NT (the latter three particularly for network data servers and server-side processing), and/or proprietary information access platforms such as Microsoft's WebTV or the Diva Systems video-on-demand system.

2. Processing Methodology

The present invention provides a spoken natural language interface for interrogation of remote electronic databases and retrieval of desired information. A preferred embodiment of the present invention utilizes the basic methodology outlined in the flow diagram of FIG. **4** in order to provide this interface. This methodology will now be discussed.

a. Interpreting Spoken Natural Language Requests

At step **402**, the user's spoken request for information is initially received in the form of raw (acoustic) voice data by a suitable input device, as previously discussed in connection with FIGS. **1**–**2**. At step **404** the voice data received from the user is interpreted in order to understand the user's request for information. Preferably this step includes performing speech recognition in order to extract words from the voice data, and further includes natural language parsing of those words in order to generate a structured linguistic representation of the user's request.

Speech recognition in step **404** is performed using speech recognition engine **310**. A variety of commercial quality, speech recognition engines are readily available on the market, as practitioners will know. For example, Nuance Communications offers a suite of speech recognition engines, including Nuance 6, its current flagship product, and Nuance Express, a lower cost package for entry-level applications. As one other example, IBM offers the ViaVoice speech recognition engine, including a low-cost shrink-wrapped version available through popular consumer distribution channels. Basically, a speech recognition engine processes acoustic voice data and attempts to generate a text stream of recognized words.

Typically, the speech recognition engine is provided with a vocabulary lexicon of likely words or phrases that the recognition engine can match against its analysis of acoustical signals, for purposes of a given application. Preferably, the lexicon is dynamically adjusted to reflect the current user context, as established by the preceding user inputs. For example, if a user is engaged in a dialogue with the system about movie selection, the recognition engine's vocabulary may preferably be adjusted to favor relevant words and phrases, such as a stored list of proper names for popular movie actors and directors, etc. Whereas if the current dialogue involves selection and viewing of a sports event, the engine's vocabulary might preferably be adjusted to favor a stored list of proper names for professional sports teams, etc. In addition, a speech recognition engine is provided with language models that help the engine predict the most likely interpretation of a given segment of acoustical voice data, in the current context of phonemes or words in which the segment appears. In addition, speech recognition engines often echo to the user, in more or less real-time, a transcription of the engine's best guess at what the user has said, giving the user an opportunity to confirm or reject.

In a further aspect of step **404**, natural language interpreter (or parser) **320** linguistically parses and interprets the textual output of the speech recognition engine. In a preferred embodiment of the present invention, the natural-language interpreter attempts to determine both the meaning of spoken words (semantic processing) as well as the grammar of the statement (syntactic processing), such as the Gemini Natural Language Understanding System developed by SRI International. The Gemini system is described in detail in publications entitled "Gemini: A Natural Language System for Spoken-Language Understanding" and "Interleaving Syntax and Semantics in an Efficient Bottom-Up Parser," both of which are currently available online at http://www.ai.sri.com/natural-language/projects/arpa-sls/nat-lang.html. (Copies of those publications are also included in an information disclosure statement submitted herewith, and are incorporated herein by this reference). Briefly, Gemini applies a set of syntactic and semantic grammar rules to a word string using a bottom-up parser to generate a logical form, which is a structured representation of the context-independent meaning of the string. Gemini can be used with a variety of grammars, including general English grammar as well as application-specific grammars. The Gemini parser is based on "unification grammar," meaning that grammatical categories incorporate features that can be assigned values; so that when grammatical category expressions are matched in the course of parsing or semantic interpretation, the information contained in the features is combined, and if the feature values are incompatible the match fails.

It is possible for some applications to achieve a significant reduction in speech recognition error by using the natural-language processing system to re-score recognition hypotheses. For example, the grammars defined for a language parser like Gemini may be compiled into context-free grammar that, in turn, can be used directly as language models for speech recognition engines like the Nuance recognizer. Further details on this methodology are provided in the publication "Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS" which is currently available online through http://www.ai.sri.com/natural-language/projects/arpa-sls/spnl-int.html. A copy of this publication is included in an information disclosure submitted herewith, and is incorporated herein by this reference.

In an embodiment of the present invention that may be preferable for some applications, the natural language interpreter "learns" from the past usage patterns of a particular user or of groups of users. In such an embodiment, the successfully interpreted requests of users are stored, and can then be used to enhance accuracy by comparing a current request to the stored requests, thereby allowing selection of a most probable result.

b. Constructing Navigation Queries

In step **405** request processing logic **300** identifies and selects an appropriate online data source where the desired information (in this case, current weather reports for a given city) can be found. Such selection may involve look-up in a locally stored table, or possibly dynamic searching through an online search engine, or other online search techniques. For some applications, an embodiment of the present invention may be implemented in which only access to a particular data source (such as a particular vendor's proprietary content database) is supported; in that case, step **405** may be trivial or may be eliminated entirely.

Step **406** attempts to construct a navigation query, reflecting the interpretation of step **404**. This operation is preferably performed by query construction logic **330**.

A "navigation query" means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information. In other words, a navigation query is constructed such that it includes whatever content and structure is required in order to access desired information electronically from a particular database or data source of interest.

For example, for many existing electronic databases, a navigation query can be embodied using a formal database query language such as Standard Query Language (SQL). For many databases, a navigation query can be constructed through a more user-friendly interactive front-end, such as a series of menus and/or interactive forms to be selected or filled in. SQL is a standard interactive and programming language for getting information from and updating a database. SQL is both an ANSI and an ISO standard. As is well known to practitioners, a Relational Database Management System (RDBMS), such as Microsoft's Access, Oracle's Oracle7, and Computer Associates' CA-OpenIngres, allow programmers to create, update, and administer a relational database. Practitioners of ordinary skill in the art will be thoroughly familiar with the notion of database navigation through structured query, and will be readily able to appreciate and utilize the existing data structures and navigational mechanisms for a given database, or to create such structures and mechanisms where desired.

In accordance with the present invention, the query constructed in step **406** must reflect the user's request as interpreted by the speech recognition engine and the NL parser in step **404**. In embodiments of the present invention wherein data source **110** (or **210** in the corresponding embodiment of FIG. **2**) is a structured relational database or the like, step **406** of the present invention may entail constructing an appropriate Structured Query Language (SQL) query or the like, or automatically filling out a front-end query form, series of menus or the like, as described above.

In many existing Internet (and Intranet) applications, an online electronic data source is accessible to users only through the medium of interaction with a so-called Common Gateway Interface (CGI) script. Typically the user who visits a web site of this nature must fill in the fields of an online interactive form. The online form is in turn linked to a CGI script, which transparently handles actual navigation of the associated data source and produces output for viewing by the user's web browser. In other words, direct user access to the data source is not supported, only mediated access through the form and CGI script is offered.

For applications of this nature, an advantageous embodiment of the present invention "scrapes" the scripted online site where information desired by a user may be found in order to facilitate construction of an effective navigation query. For example, suppose that a user's spoken natural language request is: "What's the weather in Miami?" After this request is received at step **402** and interpreted at step **404**, assume that step **405** determines that the desired weather information is available online through the medium of a CGI-scripted interactive form. Step **406** is then preferably carried out using the expanded process diagrammed in FIG. **5**. In particular, at sub-step **520**, query construction logic **330** electronically "scrapes" the online interactive form, meaning that query construction logic **330** automatically extracts the format and structure of input fields accepted by the online form. At sub-step **522**, a navigation query is then constructed by instantiating (filling in) the extracted input format—essentially an electronic template—in a manner reflecting the user's request for information as interpreted in step **404**. The flow of control then returns to step **407** of FIG. **4**. Ultimately, when the query thus constructed by scraping is used to navigate the online data source in step **408**, the query effectively initiates the same scripted response as if a human user had visited the online site and had typed appropriate entries into the input fields of the online form.

In the embodiment just described, scraping step **520** is preferably carried out with the assistance of an online extraction utility such as WebL. WebL is a scripting language for automating tasks on the World Wide Web. It is an imperative, interpreted language that has built-in support for common web protocols like HTTP and FTP, and popular data types like HTML and XML. WebL's implementation language is Java, and the complete source code is available from Compaq. In addition, step **520** is preferably performed dynamically when necessary—in other words, on-the-fly in response to a particular user query—but in some applications it may be possible to scrape relatively stable (unchanging) web sites of likely interest in advance and to cache the resulting template information.

It will be apparent, in light of the above teachings, that preferred embodiments of the present invention can provide a spoken natural language interface atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the linear browsing architecture or other artifacts of an existing menu/text/click navigation system. For example, users of an appropriate embodiment of the present invention for a video-on-demand application can directly speak the natural request: "Show me the movie 'Unforgiven'"—instead of walking step-by-step through a typically linear sequence of genre/title/actor/director menus, scrolling and selecting from potentially long lists on each menu, or instead of being forced to use an alphanumeric keyboard that cannot be as comfortable to hold or use as a lightweight remote control. Similarly, users of an appropriate embodiment of the present invention for a web-surfing application in accordance with the process shown in FIG. **5** can directly speak the natural request: "Show me a one-month price chart for Microsoft stock"—instead of potentially having to navigate to an appropriate web site, search for the right ticker symbol, enter/select the symbol, and specify display of the desired one-month price chart, each of those steps potentially involving manual navigation and data entry to one or more different interaction screens. (Note that these examples are offered to illustrate some of the potential benefits offered by appropriate embodiments of the present invention, and not to limit the scope of the invention in any respect.)

### c. Error Correction

Several problems can arise when attempting to perform searches based on spoken natural language input. As indicated at decision step **407** in the process of FIG. **4**, certain deficiencies may be identified during the process of query construction, before search of the data source is even attempted. For example, the user's request may fail to specify enough information in order to construct a navigation query that is specific enough to obtain a satisfactory search result. For example, a user might orally request "what's the weather?" whereas the national online data source identified in step **405** and scraped in step **520** might require specifying a particular city.

Additionally, certain deficiencies and problems may arise following the navigational search of the data source at step **408**, as indicated at decision step **409** in FIG. **4**. For example, with reference to a video-on-demand application, a user may wish to see the movie "Unforgiven", but perhaps the user can't recall name of the film, but knows it was directed by and starred actor Clint Eastwood. A typical video-on-demand database might indeed be expected to allow queries specifying the name of a leading actor and/or director, but in the case of this query—as in many cases—that will not be enough to narrow the search to a single film, and additional user input in some form is required.

In the event that one or more deficiencies in the user's spoken request, as processed, result in the problems described, either at step **407** or **409**, some form of error handling is in order. A straightforward, crude technique might be for the system to respond simply "input not understood/insufficient; please try again." However, that approach will likely result in frustrated users, and is not optimal or even acceptable for most applications. Instead, a preferred technique in accordance with the present invention handles such errors and deficiencies in user input at step **412**, whether detected at step **407** or step **409**, by soliciting additional input from the user in a manner taking advantage of the partial construction already performed and via user interface modalities in addition to spoken natural language ("multi-modality"). This supplemental interaction is preferably conducted through client display device **112** (**202**, in the embodiment of FIG. **2**), and may include textual, graphical, audio and/or video media. Further details and examples are provided below. Query refinement logic **340** preferably carries out step **412**. The additional input received from the user is fed into and augments interpreting step **404**, and query construction step **406** is likewise repeated with the benefit of the augmented interpretation. These operations, and subsequent navigation step **408**, are preferably repeated until no remaining problems or deficiencies are identified at decision points **407** or **409**. Further details and examples for this query refinement process are provided immediately below.

Consider again the example in which the user of a video-on-demand application wishes to see "Unforgiven" but can only recall that it was directed by and starred Clint Eastwood. First, it bears noting that using a prior art navigational interface, such as a conventional menu interface, will likely be relatively tedious in this case. The user can proceed through a sequence of menus, such as Genre (select "western"), Title (skip), Actor ("Clint Eastwood"), and Director ("Clint Eastwood"). In each case—especially for the last two items—the user would typically scroll and select from fairly long lists in order to enter his or her desired name, or perhaps use a relatively couch-unfriendly keypad to manually type the actor's name twice.

Using a preferred embodiment of the present invention, the user instead speaks aloud, holding remote control microphone **102**, "I want to see that movie starring and directed by Clint Eastwood. Can't remember the title." At step **402** the voice data is received. At step **404** the voice data is interpreted. At step **405** an appropriate online data source is selected (or perhaps the system is directly connected to a proprietary video-on-demand provider). At step **406** a query is automatically constructed by the query construction logic **330** specifying "Clint Eastwood" in both the actor and director fields. Step **407** detects no obvious problems, and so the query is electronically submitted and the data source is navigated at step **408**, yielding a list of several records satisfying the query (e.g., "Unforgiven", "True Crime", "Absolute Power", etc.). Step **409** detects that additional user input is needed to further refine the query in order to select a particular film for viewing.

At that point, in step **412** query refinement logic **340** might preferably generate a display for client display device **112** showing the (relatively short) list of film titles that satisfy the user's stated constraints. The user can then preferably use a relatively convenient input modality, such as buttons on the remote control, to select the desired title from the menu. In a further preferred embodiment, the first title on the list is highlighted by default, so that the user can simply press an "OK" button to choose that selection. In a further preferred feature, the user can mix input modalities by speaking a response like "I want number one on the list." Alternatively, the user can preferably say, "Let's see Unforgiven," having now been reminded of the title by the menu display.

Utilizing the user's supplemental input, request processing logic **300** iterates again through steps **404** and **406**, this time constructing a fully-specified query that specifically requests the Eastwood film "Unforgiven." Step **408** navigates the data source using that query and retrieves the desired film, which is then electronically transmitted in step **410** from network server **108** to client display device **112** via communications network **106**.

Now consider again the example in which the user of a web surfing application wants to know his or her local weather, and simply asks, "what's the weather?" At step **402** the voice data is received. At step **404** the voice data is interpreted. At step **405** an online web site providing current weather information for major cities around the world is selected. At step **406** and sub-step **520**, the online site is scraped using a WebL-style tool to extract an input template for interacting with the site. At sub-step **522**, query construction logic **330** attempts to construct a navigation query by instantiating the input template, but determines (quite rightly) that a required field—name of city—cannot be determined from the user's spoken request as interpreted in step **404**. Step **407** detects this deficiency, and in step **412** query refinement logic **340** preferably generates output for client display device **112** soliciting the necessary supplemental input. In a preferred embodiment, the output might display the name of the city where the user is located highlighted by default. The user can then simply press an "OK" button—or perhaps mix modalities by saying "yes, exactly"—to choose that selection. A preferred embodiment would further display an alphabetical scrollable menu listing other major cities, and/or invite the user to speak or select the name of the desired city.

Here again, utilizing the user's supplemental input, request processing logic **300** iterates through steps **404** and **406**. This time, in performing sub-step **520**, a cached version of the input template already scraped in the previous iteration might preferably be retrieved. In sub-step **522**, query construction logic **330** succeeds this time in instantiating the input template and constructing an effective query, since the desired city has now been clarified. Step **408** navigates the data source using that query and retrieves the desired weather information, which is then electronically transmitted in step **410** from network server **108** to client display device **112** via communications network **106**.

It is worth noting that in some instances, there may be details that are not explicitly provided by the user, but that query construction logic **330** or query refinement logic **340** may preferably deduce on their own through reasonable assumptions, rather than requiring the use to provide explicit clarification. For example, in the example previously described regarding a request for a weather report, in some applications it might be preferable for the system to simply assume that the user means a weather report for his or her home area and to retrieve that information, if the cost of doing so is not significantly greater than the cost of asking the user to clarify the query. Making such an assumption might be even more strongly justified in a preferred embodiment, as described earlier, where user histories are tracked, and where such history indicates that a particular user or group of users typically expect local information when asking for a weather forecast. At any rate, in the event such an assumption is made, if the user actually intended to request the weather for a different city, the user would then need to ask his or her question again. It will be apparent to practitioners, in light of the above teachings, that the choice of whether to program query construction logic **330** and query refinement logic **340** to make particular assumptions will typically involve trade-offs involving user convenience that can be assessed in the context of specific applications.

### 3. Open Agent Architecture (OAA®)

Open Agent Architecture™ (OAA®) is a software platform, developed by the assignee of the present invention, that enables effective, dynamic collaboration among communities of distributed electronic agents. OAA is described in greater detail in co-pending U.S. patent application Ser. No. 09/225,198, which has been incorporated herein by reference. Very briefly, the functionality of each client agent is made available to the agent community through registration of the client agent's capabilities with a facilitator. A software "wrapper" essentially surrounds the underlying application program performing the services offered by each client. The common infrastructure for constructing agents is preferably supplied by an agent library. The agent library is preferably accessible in the runtime environment of several different programming languages. The agent library preferably minimizes the effort required to construct a new system and maximizes the ease with which legacy systems can be "wrapped" and made compatible with the agent-based architecture of the present invention. When invoked, a client agent makes a connection to a facilitator, which is known as its parent facilitator. Upon connection, an agent registers with its parent facilitator a specification of the capabilities and services it can provide, using a highlevel, declarative Interagent Communication Language ("ICL") to express those capabilities. Tasks are presented to the facilitator in the form of ICL goal expressions. When a facilitator determines that the registered capabilities of one of its client agents will help satisfy a current goal or sub-goal thereof, the facilitator delegates that subgoal to the client agent in the form of an ICL request. The client agent processes the request and returns answers or information to the facilitator. In processing a request, the client agent can use ICL to request services of other agents, or utilize other infrastructure services for collaborative work. The facilitator coordinates and integrates the results received from different client agents on various sub-goals, in order to satisfy the overall goal.

OAA provides a useful software platform for building systems that integrate spoken natural language as well as other user input modalities. For example, see the above-referenced co-pending patent application, especially FIG. **13** and the corresponding discussion of a "multi-modal maps" application, and FIG. **12** and the corresponding discussion of a "unified messaging" application. Another example is the InfoWiz interactive information kiosk developed by the assignee and described in the document entitled "InfoWiz: An Animated Voice Interactive Information System" available online at http://www.ai.sri.com/~oaa/applications.html. A copy of the InfoWhiz document is provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference. A further example is the "CommandTalk" application developed by the assignee for the U.S. military, as described online at http://www.ai.sri.com/~lesaf/commandtalk.html and in the following publications, copies of which are provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference:

"CommandTalk: A Spoken-Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, DC, pp. 1–7, Association for Computational Linguistics

"The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty-Seventh Annual Meeting of the ACL", pp. 183–190, University of Maryland, College Park, Md., Association for Computational Linguistics

"Interpreting Language in Context in CommandTalk", 1999, by John Dowding and Elizabeth Owen Bratt and Sharon Goldwater, in "Communicative Agents: The Use of Natural Language in Embodied Systems", pp. 63–67, Association for Computing Machinery (ACM) Special Interest Group on Artificial Intelligence (SIGART), Seattle, Wash.

For some applications and systems, OAA can provide an advantageous platform for constructing embodiments of the present invention. For example, a representative application is now briefly presented, with reference to FIG. **6**. If the statement "show me movies starring John Wayne" is spoken into the voice input device, the voice data for this request will be sent by UI agent **650** to facilitator **600**, which in turn will ask natural language (NL) agent **620** and speech recognition agent **610** to interpret the query and return the interpretation in ICL format. The resulting ICL goal expression is then routed by the facilitator to appropriate agents—in this case, video-on-demand database agent **640**—to execute the request. Video database agent **640** preferably includes or is coupled to an appropriate embodiment of query construction logic **330** and query refinement logic **340**, and may also issue ICL requests to facilitator **600** for additional assistance—e.g., display of menus and capture of additional user input in the event that query refinement is needed—and facilitator **600** will delegate such requests to appropriate client agents in the community. When the desired video content is ultimately retrieved by video database agent **640**, UI agent **650** is invoked by facilitator **600** to display the movie.

Other spoken user requests, such as a request for the current weather in New York City or for a stock quote, would eventually lead facilitator to invoke web database agent **630** to access the desired information from an appropriate Internet site. Here again, web database agent **630** preferably includes or is coupled to an appropriate embodiment of query construction logic **330** and query refinement logic **340**, including a scraping utility such as WebL. Other spoken requests, such as a request to view recent emails or access voice mail, would lead the facilitator to invoke the appropriate email agent **660** and/or telephone agent **680**. A request to record a televised program of interest might lead facilitator **600** to invoke web database agent **630** to return televised program schedule information, and then invoke VCR controller agent **680** to program the associated VCR unit to record the desired television program at the scheduled time.

Control and connectivity embracing additional electronic home appliances (e.g., microwave oven, home surveillance system, etc.) can be integrated in comparable fashion. Indeed, an advantage of OAA-based embodiments of the present invention, that will be apparent to practitioners in light of the above teachings and in light of the teachings disclosed in the cited co-pending patent applications, is the relative ease and flexibility with which additional service agents can be plugged into the existing platform, immediately enabling the facilitator to respond dynamically to spoken natural language requests for the corresponding services.

## 4. Further Embodiments and Equivalents

While the present invention has been described in terms of several preferred embodiments, there are many alterations, permutations, and equivalents that may fall within the scope of this invention. It should also be noted that there are many alternative ways of implementing the methods and apparatuses of the present invention. It is therefore intended that the following appended claims be interpreted as including all such alterations, permutations, and equivalents as fall within the true spirit and scope of the present invention.

What is claimed is:

**1**. A method for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken request for desired information from the user;

(b) rendering an interpretation of the spoken request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a non-spoken modality different than the original request without requiring the user to request said non-spoken modality;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

**2**. The method of claim **1**, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and a linguistic parser.

**3**.The method of claim **1**, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

**4**. The method of claim **3**, wherein the step of extracting the input template includes dynamically scraping the online scripted interface.

**5**. The method of claim **1**, wherein the navigation query is constructed in the format of a database query language.

**6**. The method of claim **1**, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

**7**. The method of claim **1**, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

**8**. The method of claim **1**, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

**9**. The method of claim **8**, wherein the deficiencies include unresolved words of the spoken request.

**10**. The method of claim **8**, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

**11**.The method of claim **1**, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

**12**. The method of claim **11**, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

**13**. The method of claim **11**, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

**14**. The method of claim **1**, wherein the additional input is solicited upon receiving a user-input statement that additional information is required.

**15**. The method of claim **1**, wherein the step of soliciting the additional input includes presenting a menu to the user on the client device of the user.

**16**. The method of claim **1**, wherein the step of soliciting the additional input includes presenting a textual request for the additional input.

**17**. The method of claim **1**, wherein the step of soliciting the additional input includes an audible request for the additional input.

**18**. The method of claim **1**, wherein the step of soliciting the additional input includes presenting a list of portions of the electronic data source that match the navigational query.

**19**. The method of claim **1**, wherein additional input received from the user is at least partially speech based.

**20**. The method of claim **1**, wherein additional input received from the user includes no spoken input.

**21**. The method of claim **1**, wherein steps (d)–(e) are repeated until the navigational query is deemed adequate.

**22**. The method of claim **1**, wherein the input modality of step (d) includes selecting from a displayed option menu.

**23**. The method of claim **22**, wherein the act of selecting from the displayed option menu is performed by speaking.

**24**. The method of claim **1**, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

**25**. The method of claim **1**, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

**26**. The method of claim **1**, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

**27**. A system for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken request for desired information from the user;

(b) language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a non-spoken modality different than the original request without requiring the user to request said non-spoken modality;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

**28**. The system of claim **27**, wherein the language processing logic includes speech recognition logic and an linguistic parsing logic for deriving linguistic information.

**29**. The system of claim **27**, wherein the language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

**30**. The system of claim **29**, wherein the language processing logic dynamically scrapes the online scripted interface.

**31**. The system of claim **27**, wherein the query construction logic constructs the query in the format of a database query language.

**32**. The system of claim **27**, wherein at least a portion of the language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

**33**. The system of claim **27**, wherein at least a portion of the language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

**34**. The system of claim **27**, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

**35**. The system of claim **34**, wherein the deficiencies include unresolved words of the spoken request.

**36**. The system of claim **34**, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

**37**. The system of claim **27**, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

**38**. The system of claim **31**, wherein the deficiencies include existence of more than one date record within the data source responsive to the navigation query.

**39**. The system of claim **31**, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

**40**. The system of claim **27**, wherein the user Interaction logic displays an option menu.

**41**. The system of claim **40**, wherein the act of selecting from the displayed option menu is performed by speaking.

**42**. The system of claim **27**, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

**43**. The system of claim **27**, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

**44**. The system of claim **27**, wherein the display device receives data from the electronic data source on the network servers via a communications box.

**45**. The system of claim **27**, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

**46**. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a non-spoken modality different than the original request without requiring the user to request said non-spoken modality;

(e) a code segment that refines the navigation query, based upon the additional input;

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portions of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

**47**. The computer program of claim **46**, further comprising a code segment that derives linguistic information by using a speech recognition engine and a linguistic parser.

**48**. The computer program of claim **46**, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

**49**. The computer program of claim **48**, further comprising a code segment that dynamically scrapes the online scripted interface.

**50**. The computer program of claim **46**, wherein the navigation query is constructed in the format of a database query language.

**51**. The computer program of claim **46**, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

**52**. The compute program of claim **46**, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.

**53**. The computer program of claim **46**, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.

**54**. The computer program of claim **53**, wherein the deficiencies include unresolved words of the spoken request.

**55**. The computer program of claim **53**, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

**56**. The computer program of claim **46**, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.

**57**. The computer program of claim **56**, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

**58**. The computer program of claim **57**, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

**59**. The computer program of claim **46**, wherein code segment that solicits additional Input displays an option menu.

**60**. The computer program of claim **59**, wherein the act of selecting from the displayed option menu is performed by speaking.

**61**. The computer program of claim **46**, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.

**62**. The computer program of claim **46**, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

**63**. The computer program of claim **46**, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

**64**. The computer program of claim **46**, wherein the additional input is solicited upon receiving a user-input statement that additional information is required.

**65**. The computer program of claim **46**, wherein the code segment that solicits the additional input includes a code segment that presents a menu to the user on the client device of the user.

**66**. The computer program of claim **46**, wherein the code segment that solicits the additional input includes a code segment that presents a textual request for the additional input.

**67**. The computer program of claim **46**, wherein the code segment that solicits the additional input includes a code segment that produces an audible request for the additional input.

**68**. The computer program of claim **46**, wherein the code segment that solicits the additional input includes a code segment that presents a list of portions of the electronic data source that match the navigational query.

**69**. The computer program of claim **46**, wherein additional input received from the user is at least partially speech based.

**70**. The computer program of claim **46**, wherein additional input received from the user includes no spoken input.

**71**. The compute program of claim **46**, wherein code segments (d)–(e) are repeated until the navigational query is deemed adequate.

**72**. A method for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user; comprising the steps of:

(a) receiving a spoken natural language ("NL") request for desired information from the user;

(b) rendering an interpretation of the spoken request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a non-spoken modality different than the original request without requiring the user to request said non-spoken modality;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device,of the user.

**73**. The method of claim **72**, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and an NL parser.

**74**. The method of claim **72**, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

**75**. The method of claim **74**, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

**76**. The method of claim **72**, wherein the navigation query is constructed in the format of a database query language.

**77**. The method of claim **72**, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

**78**. The method of claim **72**, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

**79**. The method of claim **72**, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

**80**. The method of claim **79**, wherein the deficiencies include unresolved words of the spoken NL request.

**81**. The method of claim **79**, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

**82**. The method of claim **72**, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

**83**. The method of claim **82**, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

**84**. The method of claim **82**, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

**85**. The method of claim **72**, wherein the input modality of step (d) includes selecting from a displayed option menu.

**86**. The method of claim **85**, wherein the act of selecting from the displayed option menu is performed by speaking.

**87**. The method of claim **72**, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

**88**. The method of claim **72**, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

**89**. The method of claim **72**, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

**90**. A system or utilizing spoken natural language to navigate an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken natural language ("NL") request for desired information from the user;

(b) spoken language processing logic, operable to render an interpretation of the spoken natural language request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken natural language request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a non-spoken modality different than the original request without requiring the user to request said non-spoken modality;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

**91**. The system of claim **90**, wherein the spoken language processing logic includes speech recognition logic and an NL parsing logic for deriving linguistic information.

**92**. The system of claim **90**, wherein the spoken language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

**93**. The system of claim **90**, wherein the spoken language processing logic dynamically scrapes the online scripted interface.

**94**. The system of claim **90**, wherein the query construction logic constructs the query in the format of a database query language.

**95**. The system of claim **90**, wherein at least a portion of the spoken language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

**96**. The system of claim **90**, wherein at least a portion of the spoken language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

**97**. The system of claim **90**, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

**98**. The system of claim **97**, wherein the deficiencies include unresolved words of the spoken NL request.

**99**. The system of claim **97**, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

**100**. The system of claim **90**, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

**101**. The system of claim **100**, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

**102**. The system of claim **100**, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

**103**. The system of claim **100**, wherein the user interaction logic displays an option menu.

**104**. The system of claim **103**, wherein the act of selecting from the displayed option menu is performed by speaking.

**105**. The system of claim **90**, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

**106**. The system of claim **90**, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

**107**. The system of claim **90**, wherein the display device receives data from the electronic data source on the network servers via a communications box.

**108**. The system of claim **90**, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

**109**. A computer program embodied on a computer readable medium for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken natural language ("NL") request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken natural language request,

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a non-spoken modality different than the original request without requiring the user to request said non-spoken modality;

(e) a code segment that refines the navigation query, based upon the additional inputs;

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

**110**. The computer program of claim **109**, further comprising a code segment that derives linguistic information by using a speech recognition engine and an NL parser.

**111**. The computer program of claim **109**, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

**112**. The computer program of claim **111**, further comprising a code segment that dynamically scrapes the online scripted interface.

**113**. The computer program of claim **109**, wherein the navigation query is constructed in the format of a database query language.

**114**. The computer program of claim **109**, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

**115**. The computer program of claim **109**, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.

**116**. The computer program of claim **109**, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.

**117**. The computer program of claim **116**, wherein the deficiencies include unresolved words of the spoken NL request.

**118**. The computer program of claim **116**, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

**119**. The computer program of claim **109**, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.

**120**. The computer program of claim **119**, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

**121**. The computer program of claim **119**, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

**122**. The computer program of claim **109**, wherein code segment that solicits additional input displays an option menu.

**123**. The computer program of claim **122**, wherein the act of selecting from the displayed option menu is performed by speaking.

**124**. The computer program of claim **109**, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.

**125**. The computer program of claim **109**, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

**126**. The computer program of claim **109**, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

**127**. A method for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken natural language ("NL") request for desired information from the user;

(b) rendering an interpretation of the spoken request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a non-spoken modality different than the original request, in accordance with results generated from said at least part of a navigation query;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

**128**. The method of claim **127**, wherein the input modality of step (d) includes selecting from a displayed option menu.

**129**. The method of claim **128**, wherein the act of selecting from the displayed option menu is performed by speaking.

**130**. A method for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken natural language ("NL") request for desired information from the user;

(b) rendering an interpretation of the spoken request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a non-spoken modality different than the original request, in response to one or more deficiencies encountered during the step of constructing said at least part of a navigation query;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

**131**. The method of claim **130**, wherein the input modality of step (d) includes selecting from a displayed option menu.

**132**. The method of claim **131**, wherein the act of selecting from the displayed option menu is performed by speaking.

* * * * *

2530 U.S. PTO
09/524095
03/13/00

| ISSUE CLASSIFICATION | |
|---|---|
| Class | Subclass |
| 709 | 218 |

PATENT NUMBER
6742021

# U.S. UTILITY Patent Application

O.I.P.E.

PATENT DATE MAY 2 5 2004

| APPLICATION NO. | CONT/PRIOR | CLASS | SUBCLASS | ART UNIT | EXAMINER |
|---|---|---|---|---|---|
| 09/524095 | D | 709 | | ~~2141~~ 2621 | |

APPLICANTS
Christine Halverson
Luc Julia
Dimitris Voutsas
Adam Cheyer

TITLE
Navigating network-based electronic information using spoken natural language input with multimodal error feedback

APPLICANT(S):

PTO-2040
12/99

## ISSUING CLASSIFICATION

| ORIGINAL | | CROSS REFERENCE(S) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CLASS | SUBCLASS | CLASS | SUBCLASS (ONE SUBCLASS PER BLOCK) | | | | | |
| 709 | 218 | 707 | 5 | 4 | 102 | | | |
| INTERNATIONAL CLASSIFICATION | | | | | | | | |
| G06F 15/16 | | | | | | | | |

☐ Continued on Issue Slip Inside File Jacket

☐ **TERMINAL DISCLAIMER**

☐ The term of this patent subsequent to ____________ (date) has been disclaimed.

☐ The term of this patent shall not extend beyond the expiration date of U.S Patent. No. ____________

☐ The terminal ____ months of this patent have been disclaimed.

| DRAWINGS | | |
|---|---|---|
| Sheets Drwg. | Figs. Drwg. | Print Fig. |
| ~~6~~ 7 | ~~6~~ 7 | 4 |

(Assistant Examiner) 11/21/02 (Date)

JAMES P. TRAMMELL
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 3600
(Primary Examiner) 12/16/02 (Date)

MG
(Legal Instruments Examiner) 1-4-03 (Date)

| CLAIMS ALLOWED | |
|---|---|
| Total Claims | Print Claim for O.G. |
| 132 | 1 |

NOTICE OF ALLOWANCE MAILED
12/16/02
12/17/02

ISSUE FEE J16

| Amount Due | Date Paid |
|---|---|
| ~~640.00~~ | 3-24-03 |

ISSUE BATCH NUMBER

**WARNING:**
The information disclosed herein may be restricted. Unauthorized disclosure may be prohibited by the United States Code Title 35, Sections 122, 181 and 368. Possession outside the U.S. Patent & Trademark Office is restricted to authorized employees and contractors only.

Form PTO-436A
(Rev. 6/99)

FILED WITH: ☐ DISK (CRF) ☐ FICHE ☐ CD-ROM
(Attached in pocket on right inside flap)

FORMAL DRAWINGS

(FACE)

4

jc530 U.S. PTO
09/524095
03/13/00

INITIALS [illegible]

**PATENT APPLICATION**
09524095

# CONTENTS

| | | Date Received (Incl. C. of M.) or Date Mailed | | | Date Received (Incl. C. of M.) or Date Mailed |
|---|---|---|---|---|---|
| 1. | Application ________ papers. | | 42. | | |
| 2. | CFR Decl | 5/12/00 | 43. | | |
| 3. | EOT Decl Sm Entity | 8/21/00 | 44. | | |
| 4. | Pre Amdt. A | 7/5/00 | 45. | | |
| 5. | Request Status | 4/26/00 | 46. | | |
| 6. | IDS w/refs | 5/26/00 | 47. | | |
| 7. | IDS w/refs | 3/13/00 | 48. | | |
| 7½ | Prel Amdt B | Sep 20, 2000 | | | |
| 8 | Power of Attorney | 12-4-00 | 49. | | |
| 9. | notice of acceptance | 12-11-00 | 50. | | |
| 10. | Rejection (3 mos) | 4/24/01 | 51. | | |
| 11. | Revocation/ Power of Attorney | 4-16-01 | 52. | | |
| 12. | notice of acceptance | 4-25-01 | 53. | | |
| 13. | Supple. [illegible] I.D.S. | 4-30-01 | 54. | | |
| 14. | Chg. of Address | 6-11-01 | 55. | | |
| 15. | Ext. of time ① | 9-21-01 | 56. | | |
| 16. | Power of Attorney / Change of address | 9-21-01 | 57. | | |
| 17. | notice of acceptance | 9-26-01 | 58. | | |
| 18. | Req. for reconsideration | 9-21-01 | 59. | | |
| 19. | CFR | 10-5-01 | 60. | | |
| 20. | Final Rejection 3 months | 10/10/01 | 61. | | |
| 21. | interview summary | 1-15-02 | 62. | | |
| 22. | Req For Reconsideration AE | 1-10-02 | 63. | | |
| 23. | Advisory Action | 2-19-02 | 64. | | |
| 24. | Supp. Advisory Action | 4-3-02 | 65. | | |
| 25. | Ext. of time ③ | 4-10-02 | 66. | | |
| 26. | Req. for RCE | 4-10-02 | 67. | | |
| 27. | Prel. Amdt C | 4-10-02 | 68. | | |
| 28. | Rejection (3 mos) | 5-7-02 | 69. | | |
| 29. | Change of address | 6-25-02 | 70. | | |
| 30. | Response Amdt D | 08/05/02 | 71. | | |
| 31. | Allowance | 12/16/02 | 72. | | |
| 32. | Letter RE: Comment | 3-24-03 | 73. | | |
| 33. | IDS | 8/13/02 | 74. | | |
| 34. | New Drawings | 3/24/03 | 75. | | |
| 35. | Final Drawings 1 (sheets) set | 3/24/03 | 76. | | |
| 36. | QUERY 3-15-04 | | 77. | | |
| 37. | | | 78. | | |
| 38. | | | 79. | | |
| 39. | | | 80. | | |
| 40 | | | 81. | | |
| | | | 82. | | |

9/21
10-1-01
2-9-02
4.2
4/4
12/13
4-25-03
3/9/04

(LEFT OUTSIDE)

PATENT APPLICATION SERIAL NO. ____________________

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

```
03/29/2000 SDAVIS   00000028 500384    09524095

01 FC:101         690.00 CH
02 FC:103         630.00 CH
```

PTO-1556
(5/87)

*U.S. GPO: 1998-459-082/19144

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

| | | | | |
|---|---|---|---|---|
| **SERIAL NUMBER**<br>09/524,095 | **FILING DATE**<br>03/13/2000<br>**RULE** _ | **CLASS**<br>709 | **GROUP ART UNIT**<br>2758 | **ATTORNEY DOCKET NO.**<br>SRI1P037 |

**APPLICANTS**

Christine Halverson, San Jose, CA ;
Luc Julia, Menlo Park, CA ;
Dimitris Voutsas, Thessaloniki, GREECE;
Aden J. Cheyer, Palo Alto, CA ;

**** CONTINUING DATA ***************************

THIS APPLICATION IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,719 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,720 03/17/1999

**** FOREIGN APPLICATIONS **********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** 05/12/2000

** SMALL ENTITY **

| | | | | | |
|---|---|---|---|---|---|
| Foreign Priority claimed | ☐ yes ☐ no | **STATE OR COUNTRY**<br>CA | **SHEETS DRAWING**<br>7 | **TOTAL CLAIMS**<br>55 | **INDEPENDENT CLAIMS**<br>3 |
| 35 USC 119 (a-d) conditions met | ☐ yes ☐ no ☐ Met after Allowance | | | | |
| Verified and Acknowledged | Examiner's Signature Initials | | | | |

**ADDRESS**

Hickman Stephens Coleman & Hughes LLP
PO Box 52037
Palo Alto ,CA 94303-0746

**TITLE**

Navigating network-based electronic information using spoken ~~natural language input with multimodal error feedback~~ input with multimodal error feedback

| | | |
|---|---|---|
| **FILING FEE RECEIVED**<br>1529 | FEES: Authority has been given in Paper No. _________ to charge/credit DEPOSIT ACCOUNT No. _________ for following: | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other _______________<br>☐ Credit |

Page 1 of 1

**FILE COPY**

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

**CONFIRMATION NO. 6294**

| **SERIAL NUMBER** | **FILING DATE** | **CLASS** | **GROUP ART UNIT** | **ATTORNEY DOCKET NO.** |
|---|---|---|---|---|
| 09/524,095 | 03/13/2000<br>**RULE** | 709 | 2155 | SRI1P037 |

**APPLICANTS**

Christine Halverson, San Jose, CA;
Luc Julia, Menlo Park, CA;
Dimitris Voutsas, Thessaloniki, GREECE;
Adam J. Cheyer, Palo Alto, CA;

** **CONTINUING DATA** *************************

THIS APPLICATION IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
AND CLAIMS BENEFIT OF 60/124,719 03/17/1999
AND CLAIMS BENEFIT OF 60/124,720 03/17/1999

** **FOREIGN APPLICATIONS** ********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** SMALL ENTITY **
** 05/12/2000

| | | **STATE OR COUNTRY** | **SHEETS DRAWING** | **TOTAL CLAIMS** | **INDEPENDENT CLAIMS** |
|---|---|---|---|---|---|
| Foreign Priority claimed | ☐ yes ☐ no | CA | 7 | 55 | 3 |
| 35 USC 119 (a-d) conditions met | ☐ yes ☐ no ☐ Met after Allowance | | | | |
| Verified and Acknowledged | Examiner's Signature Initials | | | | |

**ADDRESS**

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY ,NJ 07702

**TITLE**

Navigating network-based electronic information using spoken natural language input with multimodal error feedback

| | | |
|---|---|---|
| **FILING FEE RECEIVED**<br>2141 | FEES: Authority has been given in Paper No. _________ to charge/credit DEPOSIT ACCOUNT No. _________ for following: | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other ______________<br>☐ Credit |

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

* BI BDATASHEET*
Bib Data Sheet

**CONFIRMATION NO. 6294**

| **SERIAL NUMBER** | **FILING DATE** | **CLASS** | **GROUP ART UNIT** | **ATTORNEY DOCKET NO.** |
|---|---|---|---|---|
| 09/524,095 | 03/13/2000<br>**RULE** | 709 | 2155 | SRI1P037 |

**APPLICANTS**

Christine Halverson, San Jose, CA;
Luc Julia, Menlo Park, CA;
Dimitris Voutsas, Thessaloniki, GREECE;
Aden J. Cheyer, Palo Alto, CA;

** **CONTINUING DATA** *************************

THIS APPLICATION IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
AND CLAIMS BENEFIT OF 60/124,719 03/17/1999
AND CLAIMS BENEFIT OF 60/124,720 03/17/1999

** **FOREIGN APPLICATIONS** ********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** SMALL ENTITY **
** 05/12/2000

| | **STATE OR COUNTRY** | **SHEETS DRAWING** | **TOTAL CLAIMS** | **INDEPENDENT CLAIMS** |
|---|---|---|---|---|
| Foreign Priority claimed ☐ yes ☐ no<br>35 USC 119 (a-d) conditions met ☐ yes ☐ no ☐ Met after Allowance<br>Verified and Acknowledged ______ Examiner's Signature ______ Initials | CA | 7 | 55. | 3 |

**ADDRESS**

THOMASON, MOSER & PATTERSON, LLP
595 SHREWSBURY AVENUE
SUITE 100
SHREWSBURY , NJ 07702

**TITLE**

Navigating network-based electronic information using spoken natural language input with multimodal error feedback

| | | |
|---|---|---|
| **FILING FEE RECEIVED**<br>2141 | FEES: Authority has been given in Paper No. _________ to charge/credit DEPOSIT ACCOUNT No. _________ for following: | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other ________________<br>☐ Credit |

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

**CONFIRMATION NO. 6294**

| | | | | |
|---|---|---|---|---|
| **SERIAL NUMBER**<br>09/524,095 | **FILING DATE**<br>03/13/2000<br>**RULE** | **CLASS**<br>709 | **GROUP ART UNIT**<br>2158 | **ATTORNEY DOCKET NO.**<br>SRI1P037 |

**APPLICANTS**

Christine Halverson, San Jose, CA;
Luc Julia, Menlo Park, CA;
Dimitris Voutsas, Thessaloniki, GREECE;
Adam J. Cheyer, Palo Alto, CA;

**** CONTINUING DATA ***************************

THIS APPLICATION IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
AND CLAIMS BENEFIT OF 60/124,719 03/17/1999
AND CLAIMS BENEFIT OF 60/124,720 03/17/1999

**** FOREIGN APPLICATIONS **********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** SMALL ENTITY **
** 05/12/2000

| | | | | |
|---|---|---|---|---|
| Foreign Priority claimed ☐ yes ☒ no<br>35 USC 119 (a-d) conditions met ☐ yes ☒ no ☐ Met after Allowance<br>Verified and Acknowledged ______ Examiner's Signature ______ Initials | **STATE OR COUNTRY**<br>CA | **SHEETS DRAWING**<br>7 | **TOTAL CLAIMS**<br>55 | **INDEPENDENT CLAIMS**<br>3 |

**ADDRESS**

PERKINS COIE LLP
101 JEFFERSON DRIVE
MENLO PARK ,CA 94025-1114

**TITLE**

Navigating network-based electronic information using spoken natural language input with multimodal error feedback

| | | |
|---|---|---|
| **FILING FEE RECEIVED**<br>2141 | FEES: Authority has been given in Paper No. ______ to charge/credit DEPOSIT ACCOUNT No. ______ for following: | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other ______<br>☐ Credit |

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

**CONFIRMATION NO. 6294**

| **SERIAL NUMBER** | **FILING DATE** | **CLASS** | **GROUP ART UNIT** | **ATTORNEY DOCKET NO.** |
|---|---|---|---|---|
| 09/524,095 | 03/13/2000<br>**RULE** | 709 | 2758 | SRI1P037 |

**APPLICANTS**

Christine Halverson, San Jose, CA;
Luc Julia, Menlo Park, CA;
Dimitris Voutsas, Thessaloniki, GREECE;
Aden J. Cheyer, Palo Alto, CA;

** **CONTINUING DATA** *************************

THIS APPLICATION IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,719 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,720 03/17/1999

** **FOREIGN APPLICATIONS** **********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** 05/12/2000

** SMALL ENTITY **

| | **STATE OR COUNTRY** | **SHEETS DRAWING** | **TOTAL CLAIMS** | **INDEPENDENT CLAIMS** |
|---|---|---|---|---|
| Foreign Priority claimed ☐ yes ☐ no<br>35 USC 119 (a-d) conditions met ☐ yes ☐ no ☐ Met after Allowance<br>Verified and Acknowledged ______ Examiner's Signature ______ Initials | CA | 7 | 55 | 3 |

**ADDRESS**

~~KEVIN J. ZILKA~~ C. Douglas McDonald, Esq.
CARLTON FIELDS, ~~P.A.~~ et al.
P.O. BOX ~~721030~~ 3239
~~SAN JOSE ,CA 95172-1030~~ Tampa, FL 33601-3239

**TITLE**

Navigating network-based electronic information using spoken natural language input with multimodal error feedback

| **FILING FEE RECEIVED** | FEES: Authority has been given in Paper | |
|---|---|---|
| 2141 | No. ________ to charge/credit DEPOSIT ACCOUNT<br>No. ________ for following: | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other ____________<br>☐ Credit |

UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
WASHINGTON, D.C. 20231
www.uspto.gov

Bib Data Sheet

| SERIAL NUMBER | FILING DATE | CLASS | GROUP ART UNIT | ATTORNEY DOCKET NO. |
|---|---|---|---|---|
| 09/524,095 | 03/13/2000<br>RULE _ | 709 | 2758 | SRI1P037 |

**APPLICANTS**

Christine Halverson, San Jose, CA ;
Luc Julia, Menlo Park, CA ;
Dimitris Voutsas, Thessaloniki, GREECE;
Aden J. Cheyer, Palo Alto, CA ;

**** CONTINUING DATA ***************************

THIS APPLICATION IS A CIP OF 09/225,198 01/05/1999
WHICH CLAIMS BENEFIT OF 60/124,718 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,719 03/17/1999
WHICH CLAIMS BENEFIT OF 60/124,720 03/17/1999

**** FOREIGN APPLICATIONS ***********************

**IF REQUIRED, FOREIGN FILING LICENSE GRANTED** ** 05/12/2000

** SMALL ENTITY **

Foreign Priority claimed ☐ yes ☒ no
35 USC 119 (a-d) conditions met ☐ yes ☒ no ☐ Met after Allowance
Verified and Acknowledged ______ Examiner's Signature ______ Initials

| STATE OR COUNTRY | SHEETS DRAWING | TOTAL CLAIMS | INDEPENDENT CLAIMS |
|---|---|---|---|
| CA | 7 | 55 | 3 |

**ADDRESS**

I. KEITH STEPHENS
CARLTON, FIELDS, WARD, EMMANUEL, SMITH & CUTLER
P.O. BOX 3239
TAMPA ,FL 33601-3239

**TITLE**

Navigating network-based electronic information using spoken ~~natural language input with multimodal error feedback~~ Input with multimodal error feedback

| FILING FEE RECEIVED | FEES: Authority has been given in Paper No. _________ to charge/credit DEPOSIT ACCOUNT No. _________ for following: | |
|---|---|---|
| 2141 | | ☐ All Fees<br>☐ 1.16 Fees ( Filing )<br>☐ 1.17 Fees ( Processing Ext. of time )<br>☐ 1.18 Fees ( Issue )<br>☐ Other ___________<br>☐ Credit |

3-15-00

A

# IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF EXPRESS MAILING

This transmittal and the documents and/or fees itemized hereon and attached hereto have been deposited as "Express Mail Post Office to Addressee" in accordance with 37 CFR §1.10 with Mailing Label Number EL357581014US.

Attorney Docket No.: SRI1P037

First Named Inventor:
HALVERSON, Christine

## UTILITY PATENT APPLICATION TRANSMITTAL (37 CFR. § 1.53(b))
(Continuation, Divisional or Continuation-in-part application)

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

☐ Duplicate for fee processing

Sir: This is a request for filing a patent application under 37 CFR. § 1.53(b) in the name of inventors:
**Christine Halverson**

For: **NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK**

This application is a ☐ Continuation ☐ Divisional ☒ Continuation-in-part

of prior Application No.: **09/225,198,** from which priority under 35 U.S.C. §120 is claimed.

Application Elements:

☒ 33 Pages of Specification, Claims and Abstract
☒ 07 Sheets of Drawings
☐ Combined Declaration and Power of Attorney
☐ Newly executed (original or copy)
☐ Copy from a prior application (37 CFR 1.63(d) for a continuation or divisional). The entire disclosure of the prior application from which a copy of the declaration is herein supplied is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
☐ Deletion of inventors Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).

Accompanying Application Parts:

☐ Assignment and Assignment Recordation Cover Sheet (recording fee of $40.00 enclosed)
☐ Power of Attorney
☐ 37 CFR 3.73(b) Statement by Assignee
☒ Information Disclosure Statement with Form PTO-1449 ☒ Copies of IDS Citations

(Revised 12/97, Pat App Trans 53(b) ContDivCIP)

☐ Preliminary Amendment (*New claims numbered after highest original claim in prior application.*)

☒ Return Receipt Postcard

☒ Small Entity Statement(s) ☒ Statement filed in prior application. Status still proper and desired.

☐ Other:

Claim For Foreign Priority

☐ Priority of ________ Application No. __________ filed on ___________ is claimed under 35 U.S.C. § 119.

☐ The certified copy has been filed in prior application U.S. Application No. _______

☐ The certified copy will follow.

Extension of Time for Prior Pending Application

☐ A Petition for Extension of Time is being concurrently filed in the prior pending application. A copy of the Petition for Extension of Time is attached.

Amendments

☐ Amend the specification by inserting before the first line the sentence: "This is a

☐ Continuation ☐ Continuation-in-part ☐ Divisional

application of copending prior

☐ Application No._________ filed on _____________,

☐ International Application _________ filed on _________ which designated the United States,

the disclosure of which is incorporated herein by reference."

☐ Cancel in this application original claims ________ of the prior application before calculating the filing fee. (*At least one original independent claim must be retained.*)

Fee Calculation (37 CFR § 1.16)

| | (Col. 1)<br>NO. FILED | (Col. 2)<br>NO. EXTRA | SMALL ENTITY<br>RATE | FEE | OR | LARGE ENTITY<br>RATE | FEE |
|---|---|---|---|---|---|---|---|
| BASIC FEE | | | $345 | $345. | OR | $690 | $ |
| TOTAL CLAIMS | _____ -20 = | _____ | x09 = | $ | OR | x18 = | $ |
| INDEP CLAIMS | _____ -03 = | _____ | x39 = | $ | OR | x78 = | $ |
| [ ] Multiple Dependent Claim Presented | | | $130 = | $ | OR | $260 = | $ |
| * If the difference in Col. 1 is less than zero, enter "0" in Col. 2. | | | Total | $ | OR | Total | $ |

☐ Check No. ______ in the amount of $_______ is enclosed.

☒ The Commissioner is authorized to charge any fees beyond the amount enclosed which may be required, or to credit any overpayment, to Deposit Account No. 50-0384 (Order No. SRI1P037 ).

General Authorization for Petition for Extension of Time (37 CFR §1.136)

☒ Applicants hereby make and generally authorize any Petitions for Extensions of Time as may be needed for any subsequent filings. The Commissioner is also authorized to charge any extension fees under 37 CFR §1.17 as may be needed to Deposit Account No. 50-0384 (Order No. SRI1P037).

☒ Please send correspondence to the following address:

HICKMAN STEPHENS COLEMAN & HUGHES, LLP
P. O. Box 52037
Palo Alto, California 94303-0746
(408) 558-9950

Customer No.::

Date: March 13, 2000

**Dominic M. Kotab**
Registration No. 42,762

# NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

## BACKGROUND OF THE INVENTION

This is a Continuation In Part of co-pending U.S. Patent Application No. 09/225,198, filed January 5, 1999, Provisional U.S. Patent Application No. 60/124,718, filed March 17, 1999, Provisional U.S. Patent Application No. 60/124,720, filed March 17, 1999, and Provisional U.S. Patent Application No. 60/124,719, filed March 17, 1999, from which applications priority is claimed and these application are incorporated herein by reference.

The present invention relates generally to the navigation of electronic data by means of spoken natural language requests, and to feedback mechanisms and methods for resolving the errors and ambiguities that may be associated with such requests.

As global electronic connectivity continues to grow, and the universe of electronic data potentially available to users continues to expand, there is a growing need for information navigation technology that allows relatively naïve users to navigate and access desired data by means of natural language input. In many of the most important markets -- including the home entertainment arena, as well as mobile computing -- spoken natural language input is highly desirable, if not ideal. As just one example, the proliferation of high-bandwidth communications infrastructure for the home entertainment market (cable, satellite, broadband) enables delivery of movies-on-demand and other interactive multimedia content to the consumer's home television set. For users to take full advantage of this content stream ultimately requires interactive navigation of content databases in a manner that is too complex for user-friendly selection by means of a traditional remote-control clicker. Allowing spoken natural language requests as the input modality for rapidly searching and accessing desired content is an important objective for a successful consumer entertainment product in a context offering a dizzying range of database content choices. As further examples, this same need to drive navigation of (and transaction with) relatively complex data warehouses using spoken natural language requests applies equally to surfing the Internet/Web or other networks for general information, multimedia content, or e-commerce transactions.

In general, the existing navigational systems for browsing electronic databases and data warehouses (search engines, menus, etc.), have been designed without navigation via spoken natural language as a specific goal. So today's world is full of existing electronic data navigation systems that do not assume browsing via natural spoken commands, but rather assume text and mouse-click inputs (or in the case of TV remote controls, even less). Simply recognizing voice commands within an extremely limited vocabulary and grammar -- the spoken equivalent of button/click input (e.g., speaking "channel 5" selects TV channel 5) -- is really not sufficient by itself to satisfy the objectives described above. In order to deliver a true "win" for users, the voice-driven front-end must accept spoken natural language input in a manner that is intuitive to users. For example, the front-end should not require learning a highly specialized command language or format. More fundamentally, the front-end must allow users to speak directly in terms of what the user ultimately wants -- e.g., "I'd like to see a Western film directed by Clint Eastwood" -- as opposed to speaking in terms of arbitrary navigation structures (e.g., hierarchical layers of menus, commands, etc.) that are essentially artifacts reflecting constraints of the pre-existing text/click navigation system. At the same time, the front-end must recognize and accommodate the reality that a stream of naïve spoken natural language input will, over time, typically present a variety of errors and/or ambiguities: e.g., garbled/unrecognized words (did the user say "Eastwood" or "Easter"?) and under-constrained requests ("Show me the Clint Eastwood movie"). An approach is needed for handling and resolving such errors and ambiguities in a rapid, user-friendly, non-frustrating manner.

What is needed is a methodology and apparatus for rapidly constructing a voice-driven front-end atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the step-by-step browsing architecture of the existing navigation system, and wherein any errors or ambiguities in user input are rapidly and conveniently resolved. The solution to this need should be compatible with the constraints of a multi-user, distributed environment such as the Internet/Web or a proprietary high-bandwidth content delivery network; a solution contemplating one-at-a-time user interactions at a single location is insufficient, for example.

## SUMMARY OF THE INVENTION

Sub a2

The present invention addresses the above needs by providing a system, method, and article of manufacture for navigating network-based electronic data sources in response to spoken NL input requests. When a spoken natural language input request is received from a user, it is interpreted, such as by using a speech recognition engine to extract speech data from acoustic voice signals, and using a natural language parser to linguistically parse the speech data. The interpretation of the spoken natural language request can be performed on a computing device locally with the user or remotely from the user. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is then transmitted to a client device of the user. If the network data source is a database, the navigation query is constructed in the format of a database query language.

Typically, errors or ambiguities emerge in the interpretation of the spoken NL request, such that the system cannot instantiate a complete, valid navigational template. This is to be expected occasionally, and one preferred aspect of the invention is the ability to handle such errors and ambiguities in relatively graceful and user-friendly manner. Instead of simply rejecting such input and defaulting to traditional input modes or simply asking the user to try again, a preferred embodiment of the present invention seeks to converge rapidly toward instantiation of a valid navigational template by soliciting additional clarification from the user as necessary, either before or after a navigation of the data source, via multimodal input, i.e., by means of menu selection or other input modalities including and in addition to spoken natural language. This clarifying, multi-modal dialogue takes advantage of whatever partial navigational information has been gleaned from the initial interpretation of the user's spoken NL request. This clarification process continues until the system converges toward an adequately instantiated navigational template, which is in turn used to navigate the network-based data and retrieve the user's desired information. The retrieved information is transmitted across the network and presented to the user on a suitable client display device.

In a further aspect of the present invention, the construction of the navigation query includes extracting an input template for an online scripted interface to the data source and using the input template to construct the navigation query. The extraction of the input template can include dynamically scraping the online scripted interface.

5

## BRIEF DESCRIPTION OF THE DRAWINGS

The invention, together with further advantages thereof, may best be understood by reference to the following description taken in conjunction with the accompanying drawings in which:

Figure 1a illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with server-side processing of requests;

Figure 1b illustrates another system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention with client-side processing of requests;

Figure 2 illustrates a system providing a spoken natural language interface for network-based information navigation, in accordance with an embodiment of the present invention for a mobile computing scenario;

Figure 3 illustrates the functional logic components of a request processing module in accordance with an embodiment of the present invention;

Figure 4 illustrates a process utilizing spoken natural language for navigating an electronic database in accordance with one embodiment of the present invention;

Figure 5 illustrates a process for constructing a navigational query for accessing an online data source via an interactive, scripted (e.g., CGI) form; and

Figure 6 illustrates an embodiment of the present invention utilizing a community of distributed, collaborating electronic agents.

## DETAILED DESCRIPTION OF THE INVENTION

### 1. System Architecture

a. Server-End Processing of Spoken Input

Figure 1a is an illustration of a data navigation system driven by spoken natural language input, in accordance with one embodiment of the present invention. As shown, a user's voice input data is captured by a voice input device 102, such as a microphone. Preferably voice input device 102 includes a button or the like that can be pressed or held-down to activate a listening mode, so that the system need not continually pay attention to, or be confused by, irrelevant background noise. In one preferred embodiment well-suited for the home entertainment setting, voice input device 102 is a portable remote control device with an integrated microphone, and the voice data is transmitted from device 102 preferably via infrared (or other wireless) link to communications box 104 (e.g., a set-top box or a similar communications device that is capable of retransmitting the raw voice data and/or processing the voice data) local to the user's environment and coupled to communications network 106. The voice data is then transmitted across network 106 to a remote server or servers 108. The voice data may preferably be transmitted in compressed digitized form, or alternatively --particularly where bandwidth constraints are significant-- in analog format (e.g., via frequency modulated transmission), in the latter case being digitized upon arrival at remote server 108.

At remote server 108, the voice data is processed by request processing logic 300 in order to understand the user's request and construct an appropriate query or request for navigation of remote data source 110, in accordance with the interpretation process exemplified in Figure 4 and Figure 5 and discussed in greater detail below. For purposes of executing this process, request processing logic 300 comprises functional modules including speech recognition engine 310, natural language (NL) parser 320, query construction logic 330, and query refinement logic 340, as shown in Figure 3. Data source 110 may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably resides on a central server or servers -- which may or may not be the same as server 108, depending on the storage and bandwidth needs of the application and the resources available to the practitioner. Data source 110 may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source 110 are navigated -- i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user -- using the processes of Figures 4 and 5 as described in greater detail below.

Once the desired information has been retrieved from data source 110, it is electronically transmitted via network 106 to the user for viewing on client display device 112. In a preferred embodiment well-suited for the home entertainment setting, display device 112 is a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device 112 is coupled to or integrated with a communications box (which is preferably the same as communications box 104, but may also be a separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network 106.

Network 106 is a two-way electronic communications network and may be embodied in electronic communication infrastructure including coaxial (cable television) lines, DSL, fiber-optic cable, traditional copper wire (twisted pair), or any other type of hardwired connection. Network 106 may also include a wireless connection such as a satellite-based connection, cellular connection, or other type of wireless connection. Network 106 may be part of the Internet and may support TCP/IP communications, or may be embodied in a proprietary network, or in any other electronic communications network infrastructure, whether packet-switched or connection-oriented. A design consideration is that network 106 preferably provide suitable bandwidth depending upon the nature of the content anticipated for the desired application.

### b. Client-End Processing of Spoken Input

Figure 1b is an illustration of a data navigation system driven by spoken natural language input, in accordance with a second embodiment of the present invention. Again, a user's voice input data is captured by a voice input device 102, such as a microphone. In the embodiment shown in Figure 1b, the voice data is transmitted from device 202 to requests processing logic 300, hosted on a local speech processor, for processing and interpretation. In the preferred embodiment illustrated in Figure 1b, the local speech processor is conveniently integrated as part of communications box 104, although implementation in a physically separate (but communicatively coupled) unit is also possible as will be readily apparent to those of skill in the art. The voice data is processed by the components of request processing logic 300 in order to understand the user's request and construct an appropriate query or request for navigation of remote data source 110, in accordance with the interpretation process exemplified in Figures 4 and 5 as discussed in greater detail below.

The resulting navigational query is then transmitted electronically across network 106 to data source 110, which preferably resides on a central server or servers 108. As in Figure 1a, data source 110 may comprise database(s), Internet/web site(s), or other electronic information repositories, and preferably may include multimedia content, such as movies or other digital video and audio content, other various forms of entertainment data, or other electronic information. The contents of data source 110 are then navigated -- i.e., the contents are accessed and searched, for retrieval of the particular information desired by the user -- preferably using the process of Figures 4 and 5 as described in greater detail below. Once the desired information has been retrieved from data source 110, it is electronically transmitted via network 106 to the user for viewing on client display device 112.

In one embodiment in accordance with Figure 1b and well-suited for the home entertainment setting, voice input device 102 is a portable remote control device with an integrated microphone, and the voice data is transmitted from device 102 preferably via infrared (or other wireless) link to the local speech processor. The local speech processor is coupled to communications network 106, and also preferably to client display device 112 (especially for purposes of query refinement transmissions, as discussed below in connection with Figure 4, step 412), and preferably may be integrated within or coupled to communications box 104. In addition, especially for purposes of a home entertainment application, display device 112 is preferably a television monitor or similar audiovisual entertainment device, typically in stationary position for comfortable viewing by users. In addition, in such preferred embodiment, display device 112 is coupled to a communications box (which is preferably the same as communications box 104, but may also be a physically separate unit) for receiving and decoding/formatting the desired electronic information that is received across communications network 106.

Design considerations favoring server-side processing and interpretation of spoken input requests, as exemplified in Figure 1a, include minimizing the need to distribute costly computational hardware and software to all client users in order to perform speech and language processing. Design considerations favoring client-side processing, as exemplified in Figure 1b, include minimizing the quantity of data sent upstream across the network from each client, as the speech recognition is performed before transmission across the network and only the query data and/or request needs to be sent, thus reducing the upstream bandwidth requirements.

c. Mobile Client Embodiment

A mobile computing embodiment of the present invention may be implemented by practitioners as a variation on the embodiments of either Figure 1a or Figure 1b. For example, as depicted in Figure 2, a mobile variation in accordance with the server-side processing architecture illustrated in Figure 1a may be implemented by replacing voice input device 102, communications box 104, and client display device 112, with an integrated, mobile, information appliance 202 such as a cellular telephone or wireless personal digital assistant (wireless PDA). Mobile information appliance 202 essentially performs the functions of the replaced components. Thus, mobile information appliance 202 receives spoken natural language input requests from the user in the form of voice data, and transmits that data (preferably via wireless data receiving station 204) across communications network 206 for server-side interpretation of the request, in similar fashion as described above in connection with Figure 1. Navigation of data source 210 and retrieval of desired information likewise proceeds in an analogous manner as described above. Display information transmitted electronically back to the user across network 206 is displayed for the user on the display of information appliance 202, and audio information is output through the appliance's speakers.

Practitioners will further appreciate, in light of the above teachings, that if mobile information appliance 202 is equipped with sufficient computational processing power, then a mobile variation of the client-side architecture exemplified in Figure 2 may similarly be implemented. In that case, the modules corresponding to request processing logic 300 would be embodied locally in the computational resources of mobile information appliance 202, and the logical flow of data would otherwise follow in a manner analogous to that previously described in connection with Figure 1b.

As illustrated in Figure 2, multiple users, each having their own client input device, may issue requests, simultaneously or otherwise, for navigation of data source 210. This is equally true (though not explicitly drawn) for the embodiments depicted in Figures 1a and 1b. Data source 210 (or 100), being a network accessible information resource, has typically already been constructed to support access requests from simultaneous multiple network users, as known by practitioners of ordinary skill in the art. In the case of server-side speech processing, as exemplified in Figures 1a and 2, the interpretation logic and error correction logic modules are also preferably designed and implemented to support queuing and multi-tasking of requests from multiple simultaneous network users, as will be appreciated by those of skill in the art.

It will be apparent to those skilled in the art that additional implementations, permutations and combinations of the embodiments set forth in Figures 1a, 1b, and 2 may be created without straying from the scope and spirit of the present invention. For example, practitioners will understand, in light of the above teachings and design considerations, that it is possible to divide and allocate the functional components of request processing logic 300 between client and server. For example, speech recognition -- in entirety, or perhaps just early stages such as feature extraction -- might be performed locally on the client end, perhaps to reduce bandwidth requirements, while natural language parsing and other necessary processing might be performed upstream on the server end, so that more extensive computational power need not be distributed locally to each client. In that case, corresponding portions of request processing logic 300, such as speech recognition engine 310 or portions thereof, would reside locally at the client as in Figure 1b, while other component modules would be hosted at the server end as in Figures 1a and 2.

Further, practitioners may choose to implement the each of the various embodiments described above on any number of different hardware and software computing platforms and environments and various combinations thereof, including, by way of just a few examples: a general-purpose hardware microprocessor such as the Intel Pentium series; operating system software such as Microsoft Windows/CE, Palm OS, or Apple Mac OS (particularly for client devices and client-side processing), or Unix, Linux, or Windows/NT (the latter three particularly for network data servers and server-side processing), and/or proprietary information access platforms such as Microsoft's WebTV or the Diva Systems video-on-demand system.

## 2. Processing Methodology

The present invention provides a spoken natural language interface for interrogation of remote electronic databases and retrieval of desired information. A preferred embodiment of the present invention utilizes the basic methodology outlined in the flow diagram of Figure 4 in order to provide this interface. This methodology will now be discussed.

### a. Interpreting Spoken Natural Language Requests

At step 402, the user's spoken request for information is initially received in the form of raw (acoustic) voice data by a suitable input device, as previously discussed in connection with Figures 1-2. At step 404 the voice data received from the user is interpreted in order to understand the user's request for information. Preferably this step includes performing speech recognition in order to extract words from the voice data, and further includes natural language parsing of those words in order to generate a structured linguistic representation of the user's request.

Speech recognition in step 404 is performed using speech recognition engine 310. A variety of commercial quality, speech recognition engines are readily available on the market, as practitioners will know. For example, Nuance Communications offers a suite of speech recognition engines, including Nuance 6, its current flagship product, and Nuance Express, a lower cost package for entry-level applications. As one other example, IBM offers the ViaVoice speech recognition engine, including a low-cost shrink-wrapped version available through popular consumer distribution channels. Basically, a speech recognition engine processes acoustic voice data and attempts to generate a text stream of recognized words.

Typically, the speech recognition engine is provided with a vocabulary lexicon of likely words or phrases that the recognition engine can match against its analysis of acoustical signals, for purposes of a given application. Preferably, the lexicon is dynamically adjusted to reflect the current user context, as established by the preceding user inputs. For example, if a user is engaged in a dialogue with the system about movie selection, the recognition engine's vocabulary may preferably be adjusted to favor relevant words and phrases, such as a stored list of proper names for popular movie actors and directors, etc. Whereas if the current dialogue involves selection and viewing of a sports event, the engine's vocabulary might preferably be adjusted to favor a stored list of proper names for professional sports teams, etc. In addition, a speech recognition engine is provided with language models that help the engine predict the most likely interpretation of a given segment of acoustical voice data, in the current context of phonemes or words in which the segment appears. In addition, speech recognition engines often echo to the user, in more or less real-time, a transcription of the engine's best guess at what the user has said, giving the user an opportunity to confirm or reject.

In a further aspect of step 404, natural language interpreter (or parser) 320 linguistically parses and interprets the textual output of the speech recognition engine. In a preferred embodiment of the present invention, the natural-language interpreter attempts to determine both the meaning of spoken words (semantic processing) as well as the grammar of the statement (syntactic processing), such as the Gemini Natural Language Understanding System developed by SRI International. The Gemini system is described in detail in publications entitled "Gemini: A Natural Language System for Spoken-Language Understanding" and "Interleaving Syntax and Semantics in an Efficient Bottom-Up Parser," both of which are currently available online at http://www.ai.sri.com/natural-language/projects/arpa-sls/nat-lang.html. (Copies of those publications are also included in an information disclosure statement submitted herewith, and are incorporated herein by this reference). Briefly, Gemini applies a set of syntactic and semantic grammar rules to a word string using a bottom-up parser to generate a logical form, which is a structured representation of the context-independent meaning of the string. Gemini can be used with a variety of grammars, including general English grammar as well as application-specific grammars. The Gemini parser is based on "unification grammar," meaning that grammatical categories incorporate features that can be assigned values; so that when grammatical category expressions are matched in the course of parsing or semantic interpretation, the information contained in the features is combined, and if the feature values are incompatible the match fails.

It is possible for some applications to achieve a significant reduction in speech recognition error by using the natural-language processing system to re-score recognition hypotheses. For example, the grammars defined for a language parser like Gemini may be compiled into context-free grammar that, in turn, can be used directly as language models for speech recognition engines like the Nuance recognizer. Further details on this methodology are provided in the publication "Combining Linguistic and Statistical Knowledge Sources in Natural-Language Processing for ATIS" which is currently available online through http://www.ai.sri.com/natural-language/projects/arpa-sls/spnl-int.html. A copy of this publication is included in an information disclosure submitted herewith, and is incorporated herein by this reference.

In an embodiment of the present invention that may be preferable for some applications, the natural language interpreter "learns" from the past usage patterns of a particular user or of groups of users. In such an embodiment, the successfully interpreted requests of users are stored, and can then be used to enhance accuracy by comparing a current request to the stored requests, thereby allowing selection of a most probable result.

### b. Constructing Navigation Queries

In step 405 request processing logic 300 identifies and selects an appropriate online data source where the desired information (in this case, current weather reports for a given city) can be found. Such selection may involve look-up in a locally stored table, or possibly dynamic searching through an online search engine, or other online search techniques. For some applications, an embodiment of the present invention may be implemented in which only access to a particular data source (such as a particular vendor's proprietary content database) is supported; in that case, step 405 may be trivial or may be eliminated entirely.

Step 406 attempts to construct a navigation query, reflecting the interpretation of step 404. This operation is preferably performed by query construction logic 330.

A "navigation query" means an electronic query, form, series of menu selections, or the like; being structured appropriately so as to navigate a particular data source of interest in search of desired information. In other words, a navigation query is constructed such that it includes whatever content and structure is required in order to access desired information electronically from a particular database or data source of interest.

For example, for many existing electronic databases, a navigation query can be embodied using a formal database query language such as Standard Query Language (SQL). For many databases, a navigation query can be constructed through a more user-friendly interactive front-end, such as a series of menus and/or interactive forms to be selected or filled in. SQL is a standard interactive and programming language for getting information from and updating a database. SQL is both an ANSI and an ISO standard. As is well known to practitioners, a Relational Database Management System (RDBMS), such as Microsoft's Access, Oracle's Oracle7, and Computer Associates' CA-OpenIngres, allow programmers to create, update, and administer a relational database. Practitioners of ordinary skill in the art will be thoroughly familiar with the notion of database navigation through structured query, and will be readily able to appreciate and utilize the existing data structures and navigational mechanisms for a given database, or to create such structures and mechanisms where desired.

In accordance with the present invention, the query constructed in step 406 must reflect the user's request as interpreted by the speech recognition engine and the NL parser in step 404. In embodiments of the present invention wherein data source 110 (or 210 in the corresponding embodiment of Figure 2) is a structured relational database or the like, step 406 of the present invention may entail constructing an appropriate Structured Query Language (SQL) query or the like, or automatically filling out a front-end query form, series of menus or the like, as described above.

In many existing Internet (and Intranet) applications, an online electronic data source is accessible to users only through the medium of interaction with a so-called Common Gateway Interface (CGI) script. Typically the user who visits a web site of this nature must fill in the fields of an online interactive form. The online form is in turn linked to a CGI script, which transparently handles actual navigation of the associated data source and produces output for viewing by the user's web browser. In other words, direct user access to the data source is not supported, only mediated access through the form and CGI script is offered.

For applications of this nature, an advantageous embodiment of the present invention "scrapes" the scripted online site where information desired by a user may be found in order to facilitate construction of an effective navigation query. For example, suppose that a user's spoken natural language request is: "What's the weather in Miami?" After this request is received at step 402 and interpreted at step 404, assume that step 405 determines that the desired weather information is available online through the medium of a CGI-scripted interactive form. Step 406 is then preferably carried out using the expanded process diagrammed in Figure 5. In particular, at sub-step 520, query construction logic 330 electronically "scrapes" the online interactive form, meaning that query construction logic 330 automatically extracts the format and structure of input fields accepted by the online form. At sub-step 522, a navigation query is then constructed by instantiating (filling in) the extracted input format -- essentially an electronic template -- in a manner reflecting the user's request for information as interpreted in step 404. The flow of control then returns to step 407 of Figure 4. Ultimately, when the query thus constructed by scraping is used to navigate the online data source in step 408, the query effectively initiates the same scripted response as if a human user had visited the online site and had typed appropriate entries into the input fields of the online form.

In the embodiment just described, scraping step 520 is preferably carried out with the assistance of an online extraction utility such as WebL. WebL is a scripting language for automating tasks on the World Wide Web. It is an imperative, interpreted language that has built-in support for common web protocols like HTTP and FTP, and popular data types like HTML and XML. WebL's implementation language is Java, and the complete source code is available from Compaq. In addition, step 520 is preferably performed dynamically when necessary -- in other words, on-the-fly in response to a particular user query -- but in some applications it may be possible to scrape relatively stable (unchanging) web sites of likely interest in advance and to cache the resulting template information.

It will be apparent, in light of the above teachings, that preferred embodiments of the present invention can provide a spoken natural language interface atop an existing, non-voice data navigation system, whereby users can interact by means of intuitive natural language input not strictly conforming to the linear browsing architecture or other artifacts of an existing menu/text/click navigation system. For example, users of an appropriate embodiment of the present invention for a video-on-demand application can directly speak the natural request: "Show me the movie 'Unforgiven'" -- instead of walking step-by-step through a typically linear sequence of genre/title/actor/director menus, scrolling and selecting from potentially long lists on each menu, or instead of being forced to use an alphanumeric keyboard that cannot be as comfortable to hold or use as a lightweight remote control. Similarly, users of an appropriate embodiment of the present invention for a web-surfing application in accordance with the process shown in Figure 5 can directly speak the natural request: "Show me a one-month price chart for Microsoft stock" -- instead of potentially having to navigate to an appropriate web site, search for the right ticker symbol, enter/select the symbol, and specify display of the desired one-month price chart, each of those steps potentially involving manual navigation and data entry to one or more different interaction screens. (Note that these examples are offered to illustrate some of the potential benefits offered by appropriate embodiments of the present invention, and not to limit the scope of the invention in any respect.)

### c. Error Correction

Several problems can arise when attempting to perform searches based on spoken natural language input. As indicated at decision step 407 in the process of Figure 4, certain deficiencies may be identified during the process of query construction, before search of the data source is even attempted. For example, the user's request may fail to specify enough information in order to construct a navigation query that is specific enough to obtain a satisfactory search result. For example, a user might orally request "what's the weather?" whereas the national online data source identified in step 405 and scraped in step 520 might require specifying a particular city.

Additionally, certain deficiencies and problems may arise following the navigational search of the data source at step 408, as indicated at decision step 409 in Figure 4. For example, with reference to a video-on-demand application, a user may wish to see the movie "Unforgiven", but perhaps the user can't recall name of the film, but knows it was directed by and starred actor Clint Eastwood. A typical video-on-demand database might indeed be expected to allow queries specifying the name of a leading actor and/or director, but in the case of this query -- as in many cases -- that will not be enough to narrow the search to a single film, and additional user input in some form is required.

In the event that one or more deficiencies in the user's spoken request, as processed, result in the problems described, either at step 407 or 409, some form of error handling is in order. A straightforward, crude technique might be for the system to respond simply *"input not understood / insufficient; please try again."* However, that approach will likely result in frustrated users, and is not optimal or even acceptable for most applications. Instead, a preferred technique in accordance with the present invention handles such errors and deficiencies in user input at step 412, whether detected at step 407 or step 409, by soliciting additional input from the user in a manner taking advantage of the partial construction already performed and via user interface modalities in addition to spoken natural language ("multi-modality"). This supplemental interaction is preferably conducted through client display device 112 (202, in the embodiment of Figure 2), and may include textual, graphical, audio and/or video media. Further details and examples are provided below. Query refinement logic 340 preferably carries out step 412. The additional input received from the user is fed into and augments interpreting step 404, and query construction step 406 is likewise repeated with the benefit of the augmented interpretation. These operations, and subsequent navigation step 408, are preferably repeated until no remaining problems or deficiencies are identified at decision points 407 or 409. Further details and examples for this query refinement process are provided immediately below.

Consider again the example in which the user of a video-on-demand application wishes to see "Unforgiven" but can only recall that it was directed by and starred Clint Eastwood. First, it bears noting that using a prior art navigational interface, such as a conventional menu interface, will likely be relatively tedious in this case. The user can proceed through a sequence of menus, such as Genre (select "western"), Title (skip), Actor ("Clint Eastwood"), and Director ("Clint Eastwood"). In each case --especially for the last two items -- the user would typically scroll and select from fairly long lists in order to enter his or her desired name, or perhaps use a relatively couch-unfriendly keypad to manually type the actor's name twice.

Using a preferred embodiment of the present invention, the user instead speaks aloud, holding remote control microphone 102, "I want to see that movie starring and directed by Clint Eastwood. Can't remember the title." At step 402 the voice data is received. At step 404 the voice data is interpreted. At step 405 an appropriate online data source is selected (or perhaps the system is directly connected to a proprietary video-on-demand provider). At step 406 a query is automatically constructed by the query construction logic 330 specifying "Clint Eastwood" in both the actor and director fields. Step 407 detects no obvious problems, and so the query is electronically submitted and the data source is navigated at step 408, yielding a list of several records satisfying the query (e.g., "Unforgiven", "True Crime", "Absolute Power", etc.). Step 409 detects that additional user input is needed to further refine the query in order to select a particular film for viewing.

At that point, in step 412 query refinement logic 340 might preferably generate a display for client display device 112 showing the (relatively short) list of film titles that satisfy the user's stated constraints. The user can then preferably use a relatively convenient input modality, such as buttons on the remote control, to select the desired title from the menu. In a further preferred embodiment, the first title on the list is highlighted by default, so that the user can simply press an "OK" button to choose that selection. In a further preferred feature, the user can mix input modalities by speaking a response like "I want number one on the list." Alternatively, the user can preferably say, "Let's see Unforgiven," having now been reminded of the title by the menu display.

Utilizing the user's supplemental input, request processing logic 300 iterates again through steps 404 and 406, this time constructing a fully-specified query that specifically requests the Eastwood film "Unforgiven." Step 408 navigates the data source using that query and retrieves the desired film, which is then electronically transmitted in step 410 from network server 108 to client display device 112 via communications network 106.

Now consider again the example in which the user of a web surfing application wants to know his or her local weather, and simply asks, "what's the weather?" At step 402 the voice data is received. At step 404 the voice data is interpreted. At step 405 an online web site providing current weather information for major cities around the world is selected. At step 406 and sub-step 520, the online site is scraped using a WebL-style tool to extract an input template for interacting with the site. At sub-step 522, query construction logic 330 attempts to construct a navigation query by instantiating the input template, but determines (quite rightly) that a required field -- name of city -- cannot be determined from the user's spoken request as interpreted in step 404. Step 407 detects this deficiency, and in step 412 query refinement logic 340 preferably generates output for client display device 112 soliciting the necessary supplemental input. In a preferred embodiment, the output might display the name of the city where the user is located highlighted by default. The user can then simply press an "OK" button -- or perhaps mix modalities by saying "yes, exactly" -- to choose that selection. A preferred embodiment would further display an alphabetical scrollable menu listing other major cities, and/or invite the user to speak or select the name of the desired city.

Here again, utilizing the user's supplemental input, request processing logic 300 iterates through steps 404 and 406. This time, in performing sub-step 520, a cached version of the input template already scraped in the previous iteration might preferably be retrieved. In sub-step 522, query construction logic 330 succeeds this time in instantiating the input template and constructing an effective query, since the desired city has now been clarified. Step 408 navigates the data source using that query and retrieves the desired weather information, which is then electronically transmitted in step 410 from network server 108 to client display device 112 via communications network 106.

It is worth noting that in some instances, there may be details that are not explicitly provided by the user, but that query construction logic 330 or query refinement logic 340 may preferably deduce on their own through reasonable assumptions, rather than requiring the use to provide explicit clarification. For example, in the example previously described regarding a request for a weather report, in some applications it might be preferable for the system to simply assume that the user means a weather report for his or her home area and to retrieve that information, if the cost of doing so is not significantly greater than the cost of asking the user to clarify the query. Making such an assumption might be even more strongly justified in a preferred embodiment, as described earlier, where user histories are tracked, and where such history indicates that a particular user or group of users typically expect local information when asking for a weather forecast. At any rate, in the event such an assumption is made, if the user actually intended to request the weather for a different city, the user would then need to ask his or her question again. It will be apparent to practitioners, in light of the above teachings, that the choice of whether to program query construction logic 330 and query refinement logic 340 to make make particular assumptions will typically involve trade-offs involving user conveience that can be assessed in the context of specific applications.

### 3. Open Agent Architecture (OAA®)

Open Agent Architecture™ (OAA®) is a software platform, developed by the assignee of the present invention, that enables effective, dynamic collaboration among communities of distributed electronic agents. OAA is described in greater detail in co-pending U.S. Patent Application No. 09/225,198, which has been incorporated herein by reference. Very briefly, the functionality of each client agent is made available to the agent community through registration of the client agent's capabilities with a facilitator. A software "wrapper" essentially surrounds the underlying application program performing the services offered by each client. The common infrastructure for constructing agents is preferably supplied by an *agent library*. The agent library is preferably accessible in the runtime environment of several different programming languages. The agent library preferably minimizes the effort required to construct a new system and maximizes the ease with which legacy systems can be "wrapped" and made compatible with the agent-based architecture of the present invention. When invoked, a client agent makes a connection to a facilitator, which is known as its *parent facilitator*. Upon connection, an agent registers with its parent facilitator a specification of the capabilities and services it can provide, using a high-level, declarative Interagent Communication Language (*"ICL"*) to express those capabilities. Tasks are presented to the facilitator in the form of ICL goal expressions. When a facilitator determines that the registered capabilities of one of its client agents will help satisfy a current goal or sub-goal thereof, the facilitator delegates that sub-goal to the client agent in the form of an ICL request. The client agent processes the request and returns answers or information to the facilitator. In processing a request, the client agent can use *ICL* to request services of other agents, or utilize other infrastructure services for collaborative work. The facilitator coordinates and integrates the results received from different client agents on various sub-goals, in order to satisfy the overall goal.

OAA provides a useful software platform for building systems that integrate spoken natural language as well as other user input modalities. For example, see the above-referenced co-pending patent application, especially Figure 13 and the corresponding discussion of a "multi-modal maps" application, and Figure 12 and the corresponding discussion of a "unified messaging" application. Another example is the InfoWiz interactive information kiosk developed by the assignee and described in the document entitled "InfoWiz: An Animated Voice Interactive Information System" available online at http://www.ai.sri.com/~oaa/applications.html. A copy of the InfoWhiz document is provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference. A further example is the "CommandTalk" application developed by the assignee for the U.S. military, as described online at http://www.ai.sri.com/~lesaf/commandtalk.html and in the following publications, copies of which are provided in an Information Disclosure Statement submitted herewith and incorporated herein by this reference:

- "CommandTalk: A Spoken-Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, DC, pp. 1-7, Association for Computational Linguistics

- "The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty-Seventh Annual Meeting of the ACL", pp. 183-190, University of Maryland, College Park, MD, Association for Computational Linguistics

- "Interpreting Language in Context in CommandTalk", 1999, by John Dowding and Elizabeth Owen Bratt and Sharon Goldwater, in "Communicative Agents: The Use of Natural Language in Embodied Systems", pp. 63-67, Association for Computing Machinery (ACM) Special Interest Group on Artificial Intelligence (SIGART), Seattle, WA

For some applications and systems, OAA can provide an advantageous platform for constructing embodiments of the present invention. For example, a representative application is now briefly presented, with reference to Figure 6. If the statement "show me movies starring John Wayne" is spoken into the voice input device, the voice data for this request will be sent by UI agent 650 to facilitator 600, which in turn will ask natural language (NL) agent 620 and speech recognition agent 610 to interpret the query and return the interpretation in *ICL* format. The resulting *ICL* goal expression is then routed by the facilitator to appropriate agents -- in this case, video-on-demand database agent 640 -- to execute the request. Video database agent 640 preferably includes or is coupled to an appropriate embodiment of query construction logic 330 and query refinement logic 340, and may also issue ICL

22

requests to facilitator 600 for additional assistance -- e.g., display of menus and capture of additional user input in the event that query refinement is needed -- and facilitator 600 will delegate such requests to appropriate client agents in the community. When the desired video content is ultimately retrieved by video database agent 640, UI agent 650 is invoked by facilitator 600 to display the movie.

Other spoken user requests, such as a request for the current weather in New York City or for a stock quote, would eventually lead facilitator to invoke web database agent 630 to access the desired information from an appropriate Internet site. Here again, web database agent 630 preferably includes or is coupled to an appropriate embodiment of query construction logic 330 and query refinement logic 340, including a scraping utility such as WebL. Other spoken requests, such as a request to view recent emails or access voice mail, would lead the facilitator to invoke the appropriate email agent 660 and/or telephone agent 680. A request to record a televised program of interest might lead facilitator 600 to invoke web database agent 630 to return televised program schedule information, and then invoke VCR controller agent 680 to program the associated VCR unit to record the desired television program at the scheduled time.

Control and connectivity embracing additional electronic home appliances (e.g., microwave oven, home surveillance system, etc.) can be integrated in comparable fashion. Indeed, an advantage of OAA-based embodiments of the present invention, that will be apparent to practitioners in light of the above teachings and in light of the teachings disclosed in the cited co-pending patent applications, is the relative ease and flexibility with which additional service agents can be plugged into the existing platform, immediately enabling the facilitator to respond dynamically to spoken natural language requests for the corresponding services.

## 4. Further Embodiments and Equivalents

While the present invention has been described in terms of several preferred embodiments, there are many alterations, permutations, and equivalents that may fall within the scope of this invention. It should also be noted that there are many alternative ways of implementing the methods and apparatuses of the present invention. It is therefore intended that the following appended claims be interpreted as including all such alterations, permutations, and equivalents as fall within the true spirit and scope of the present invention.

## CLAIMS

***What is claimed is:***

1. A method for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken natural language ("NL") request for desired information from the user;

(b) rendering an interpretation of the spoken natural language request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a modality different than the original request;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

2. The method of claim 1, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and an NL parser.

3. The method of claim 1, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

4. The method of claim 3, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

5. The method of claim 1, wherein the navigation query is constructed in the format of a database query language.

6. The method of claim 1, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

7. The method of claim 1, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

8. The method of claim 1, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

9. The method of claim 8, wherein the deficiencies include unresolved words of the spoken NL request.

10. The method of claim 8, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

11. The method of claim 1, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

12. The method of claim 11, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

13. The method of claim 11, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

14. The method of claim 1, wherein the input modality of step (d) includes selecting from a displayed option menu.

15. The method of claim 14, wherein the act of selecting from the displayed option menu is performed by speaking.

16. The method of claim 1, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

17. The method of claim 1, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

18. The method of claim 1, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

19. A system for utilizing spoken natural language to navigate an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken natural language ("NL") request for desired information from the user;

(b) spoken language processing logic, operable to render an interpretation of the spoken natural language request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken natural language request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a modality different than the original request;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

20. The system of claim 19, wherein the spoken language processing logic includes speech recognition logic and an NL parsing logic for deriving linguistic information.

21. The system of claim 19, wherein the spoken language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

22. The system of claim 21, wherein the spoken language processing logic dynamically scrapes the online scripted interface.

23. The system of claim 19, wherein the query construction logic constructs the query in the format of a database query language.

24. The system of claim 19, wherein at least a portion of the spoken language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

25. The system of claim 19, wherein at least a portion of the spoken language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

26. The system of claim 19, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

27. The system of claim 26, wherein the deficiencies include unresolved words of the spoken NL request.

28. The system of claim 26, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

29. The system of claim 19, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

30. The system of claim 29, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

31. The system of claim 29, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

32. The system of claim 19, wherein the user interaction logic displays an option menu.

33. The system of claim 32, wherein the act of selecting from the displayed option menu is performed by speaking.

34. The system of claim 19, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

35. The system of claim 19, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

36. The system of claim 19, wherein the display device receives data from the electronicdata source on the network servers via a communications box.

37. The system of claim 19, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

38. An computer program embodied on a computer readable medium for utilizing spoken natural language for navigating an electronic data source, the

electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken natural language ("NL") request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken natural language request;

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;

(e) a code segment that refines the navigation query, based upon the additional input;

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

39. The computer program of claim 38, further comprising a code segment that derives linguistic information by using a speech recognition engine and an NL parser.

40. The computer program of claim 38, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

41. The computer program of claim 40, further comprising a code segment that dynamically scrapes the online scripted interface.

42. The computer program of claim 38, wherein the navigation query is constructed in the format of a database query language.

~~43. The computer program of claim 38, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.~~

~~44. The computer program of claim 38, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.~~

~~45. The computer program of claim 38, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.~~

~~46. The computer program of claim 45, wherein the deficiencies include unresolved words of the spoken NL request.~~

~~47. The computer program of claim 45, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.~~

~~48. The computer program of claim 38, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.~~

~~49. The computer program of claim 48, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.~~

~~50. The computer program of claim 48, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.~~

~~51. The computer program of claim 38, wherein code segment that solicits additional input displays an option menu.~~

~~52. The computer program of claim 51, wherein the act of selecting from the displayed option menu is performed by speaking.~~

~~53. The computer program of claim 38, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.~~

~~54. The computer program of claim 38, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.~~

~~55. The computer program of claim 38, wherein the electronic data source stores multimedia content including at least one of video content and audio content.~~

~~add 8~~

~~add 15~~

add 20

# NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

## ABSTRACT OF THE INVENTION

Sub a'

~~A system, method, and article of manufacture are provided for navigating an electronic data source by means of spoken natural language. When a spoken natural language input request is received from a user, it is interpreted. Additional input is solicited from the user in a modality different than the original request and used to refine the navigation query. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources.~~

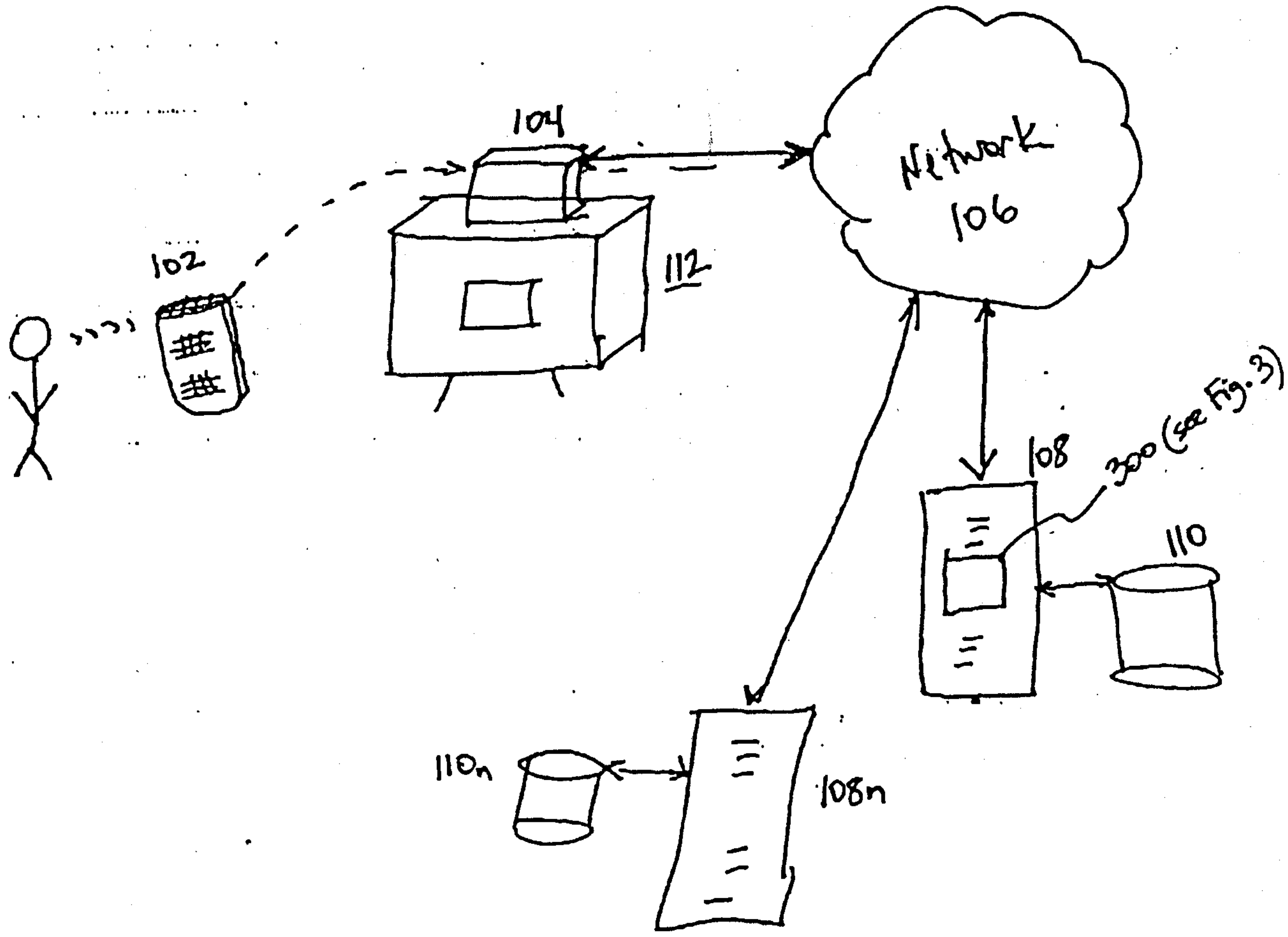


Fig. 1a

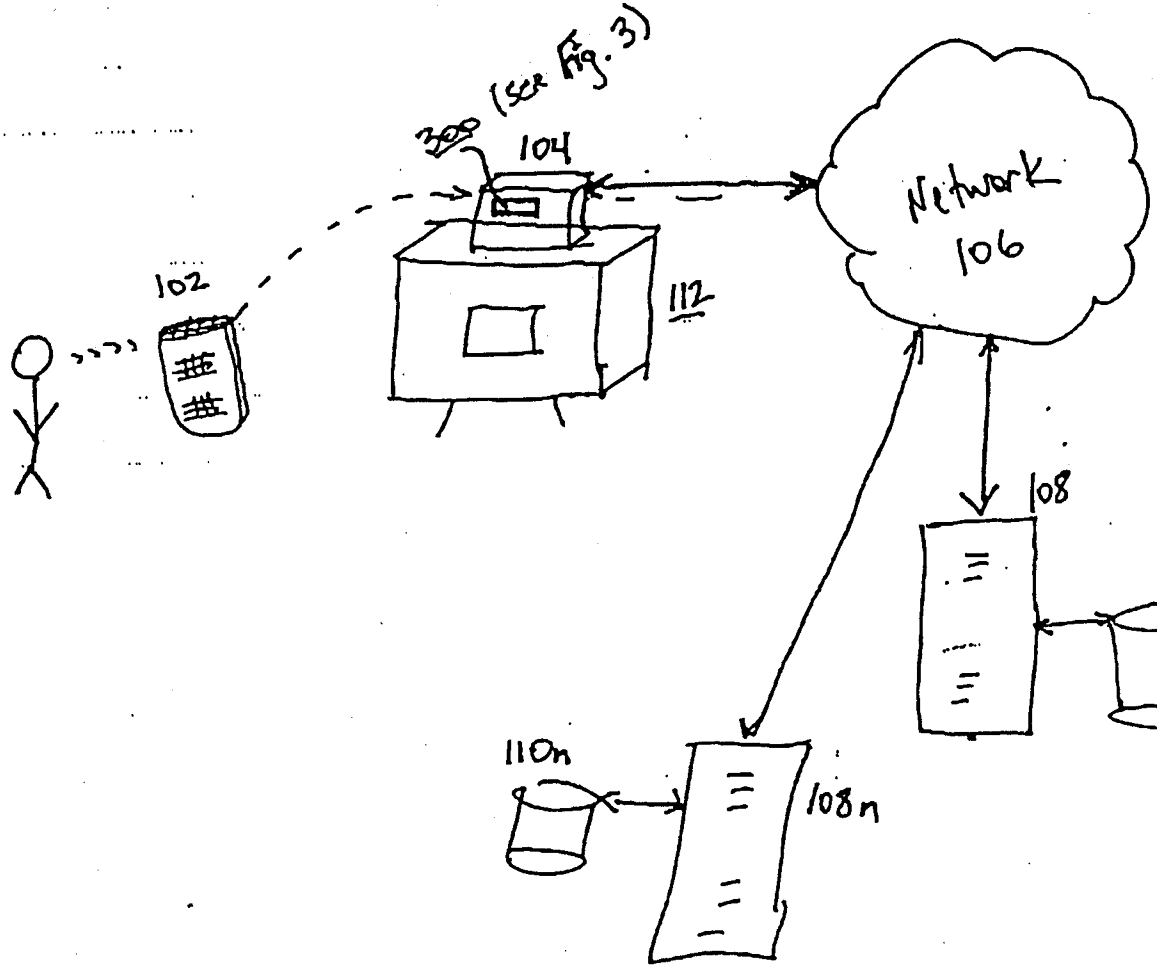


Fig. 1b

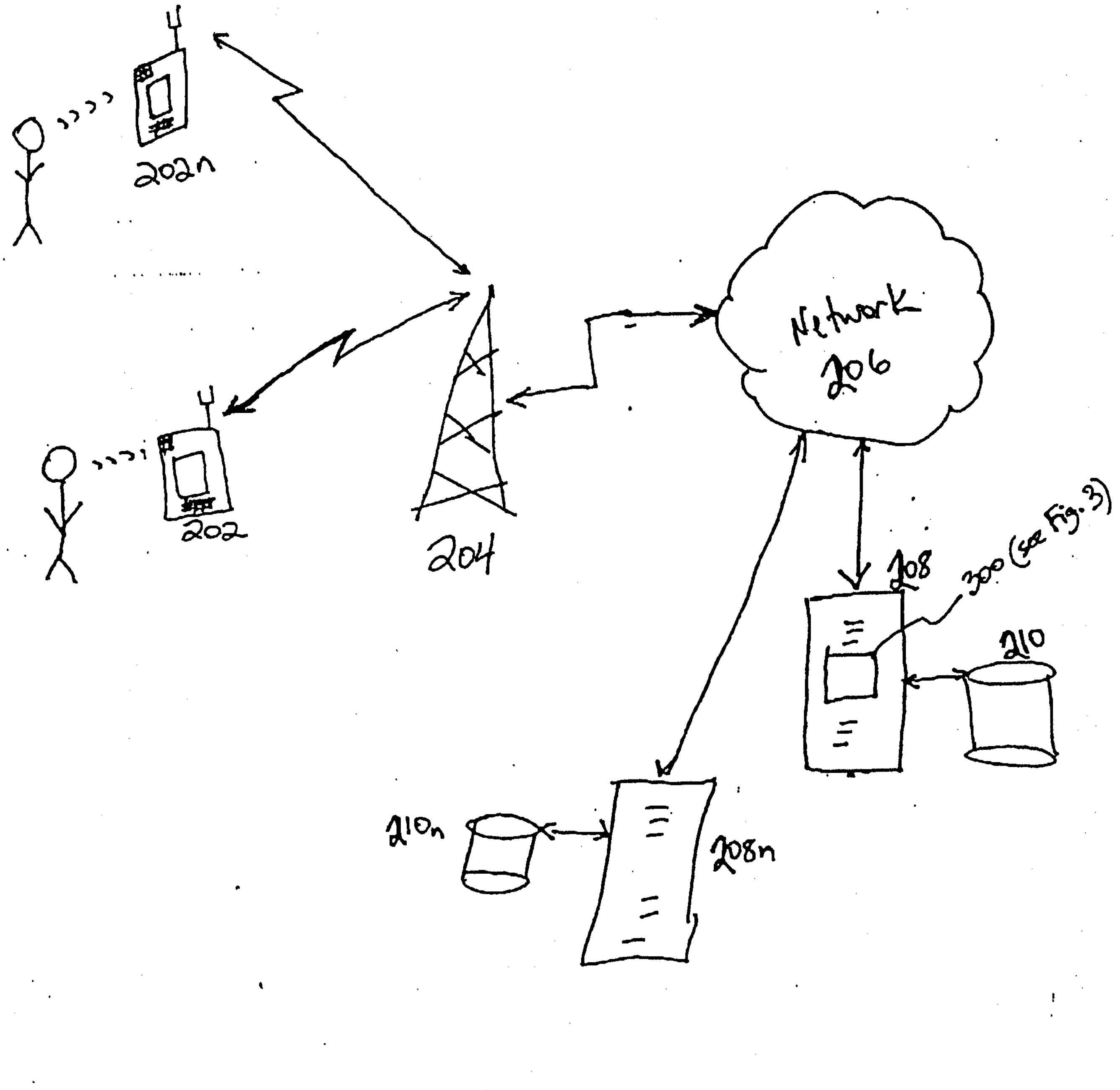


Fig. 2

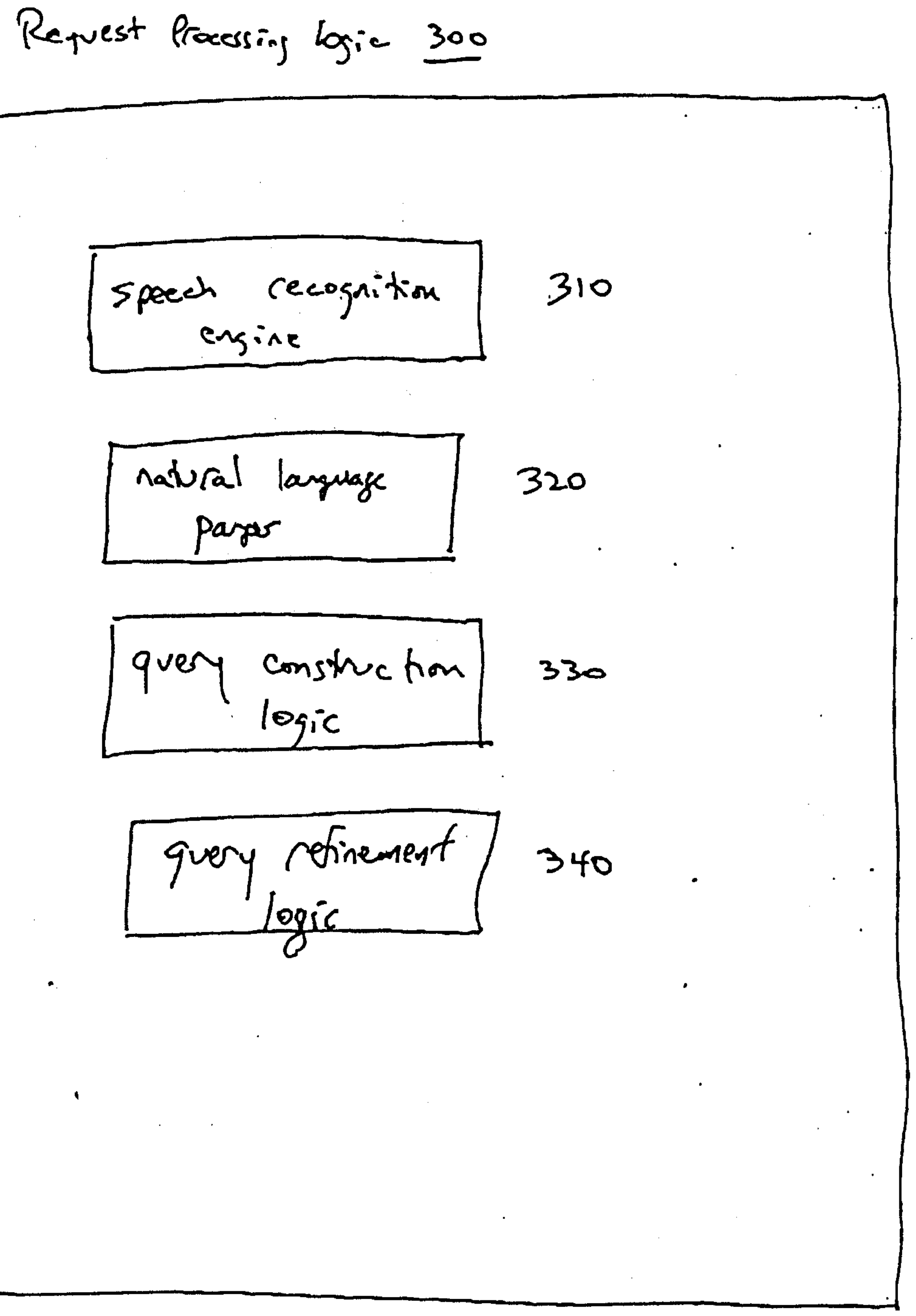

Request Processing logic 300
speech recognition engine
310
natural language parser
320
query construction logic
330
query refinement logic
340

Fig. 3

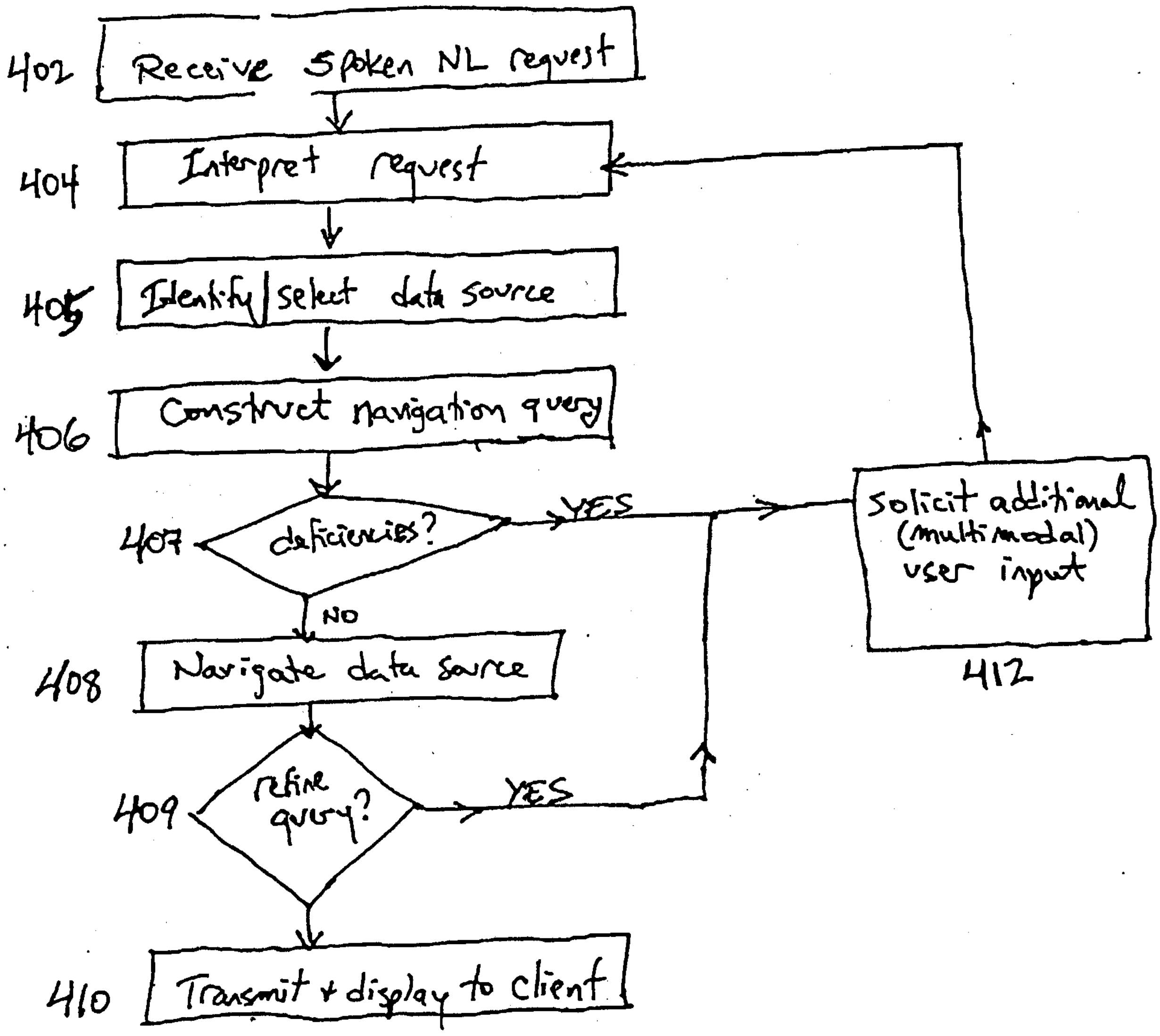

402
Receive spoken NL request
404
Interpret request
405
Identify/select data source
406
Construct navigation query
407
deficiencies?
YES
NO
Solicit additional (multimodal) user input
412
408
Navigate data source
409
refine query?
YES
410
Transmit + display to client

Fig. 4

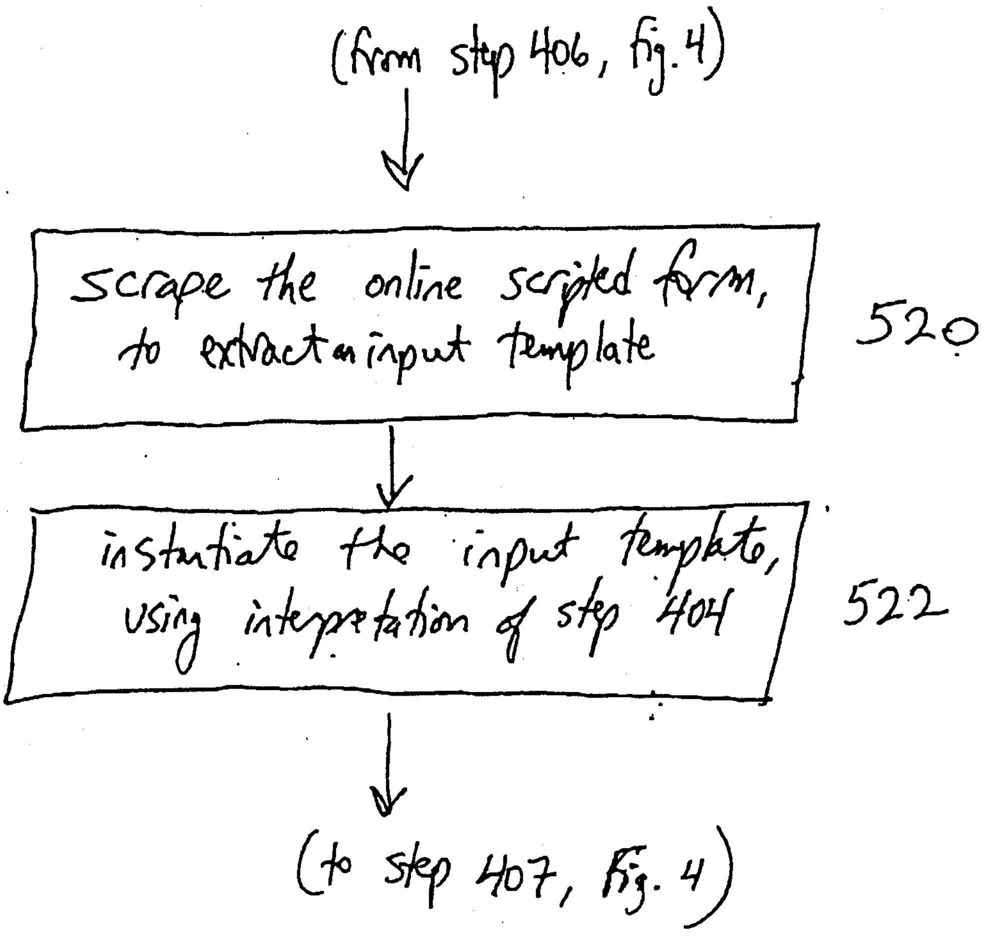

(from step 406, fig. 4)
scrape the online scripted form, to extract an input template
520
instantiate the input template, using interpretation of step 404
522
(to step 407, Fig. 4)

Fig. 5

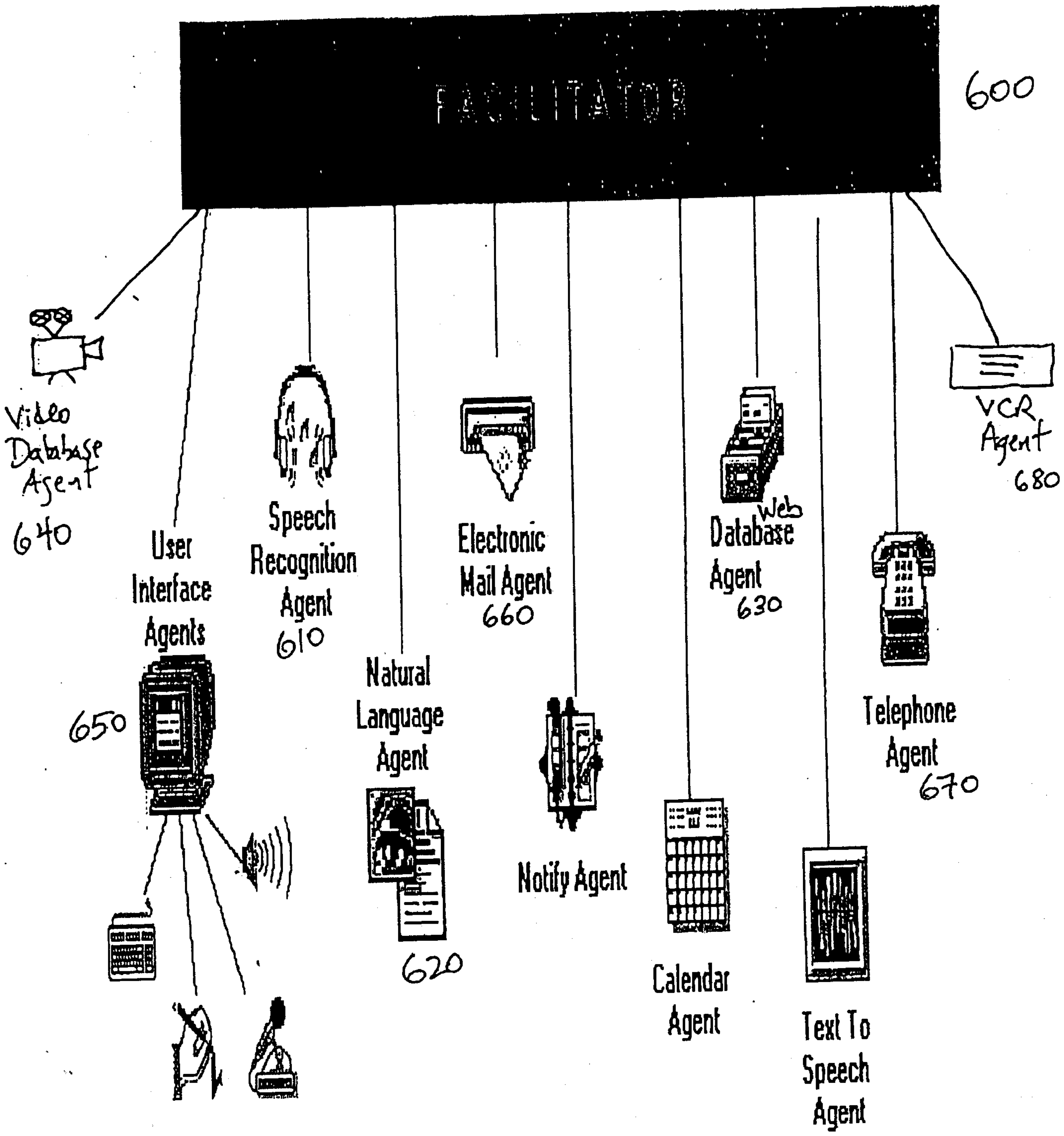

FACILITATOR
600
Video Database Agent
640
User Interface Agents
650
Speech Recognition Agent
610
Natural Language Agent
620
Electronic Mail Agent
660
Notify Agent
Calendar Agent
Web Database Agent
630
Text To Speech Agent
Telephone Agent
670
VCR Agent
680

Fig. 6

PRINT OF DRAWINGS
AS ORIGINALLY FIL

09524005 .031300

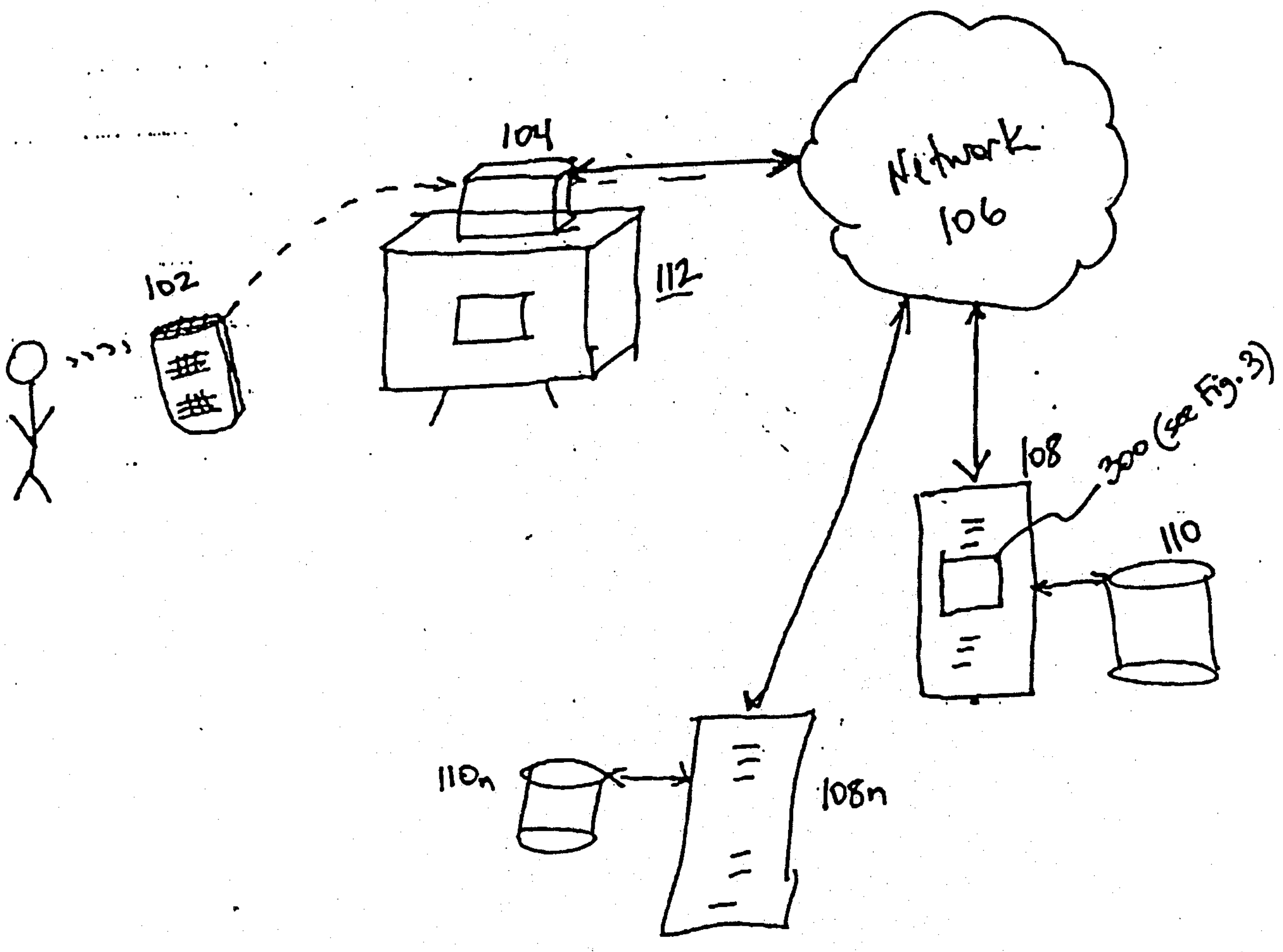


Fig. 1a

PRINT OF DRAWINGS
AS ORIGINALLY FIL

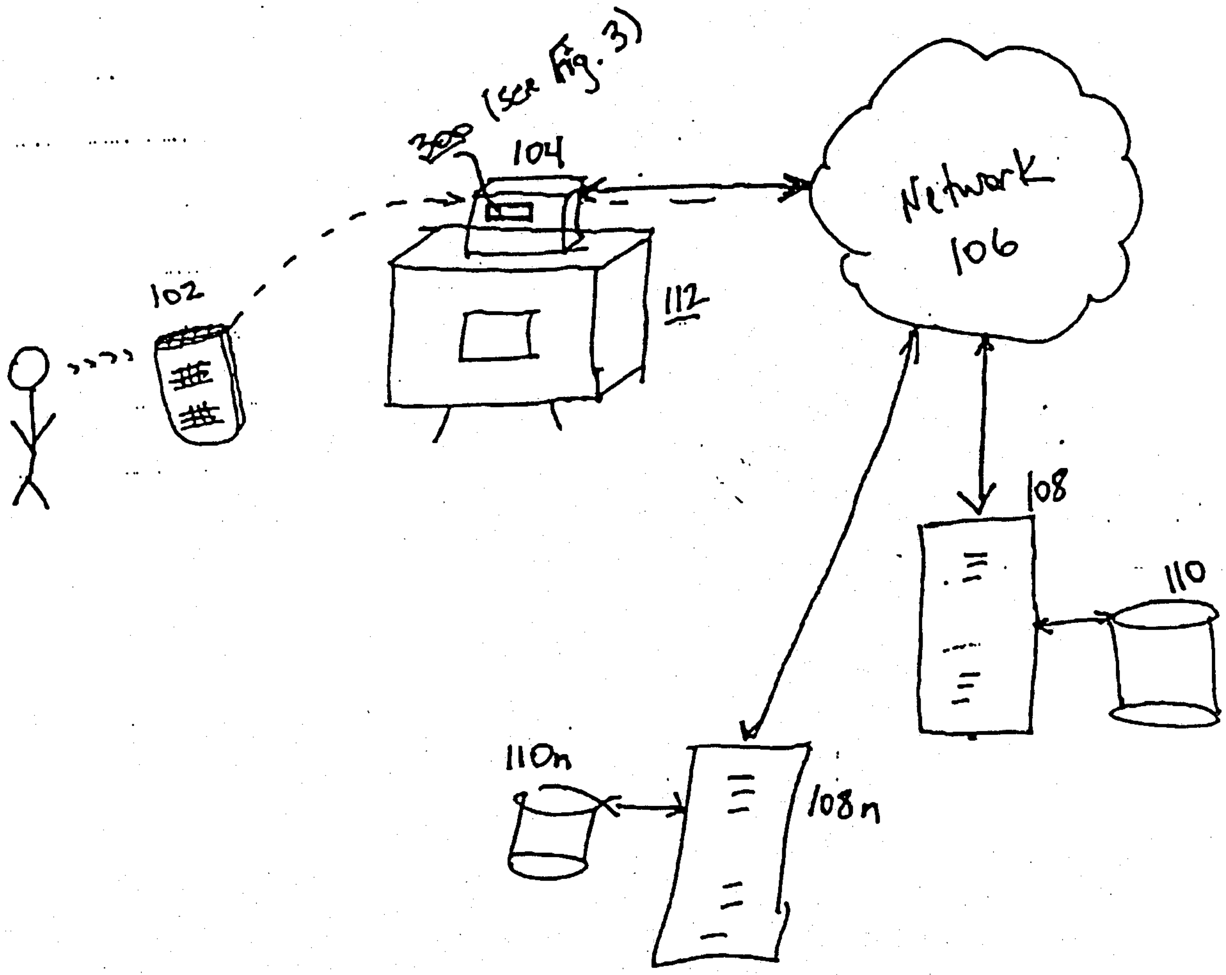


Fig. 1b

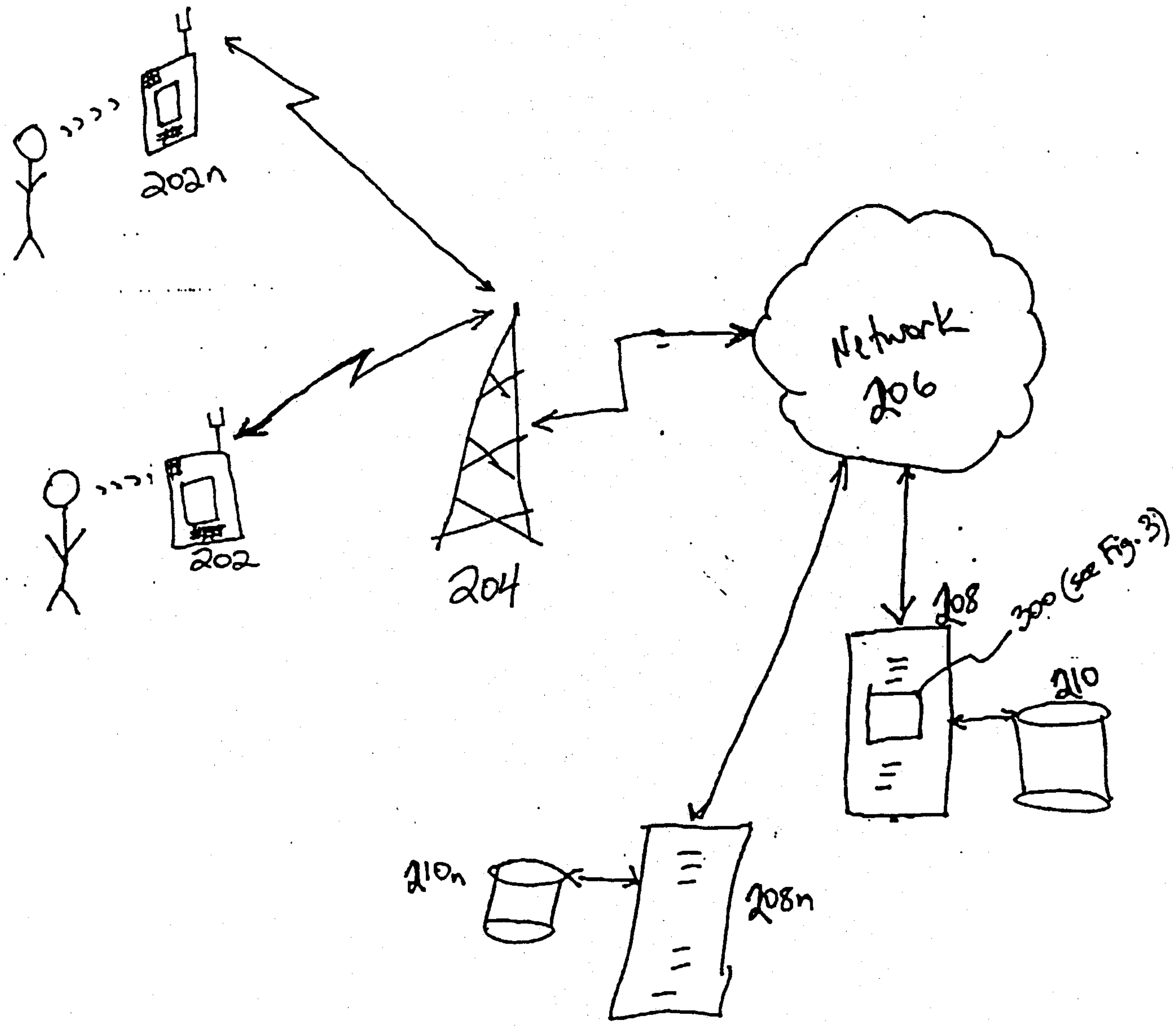


Fig. 2

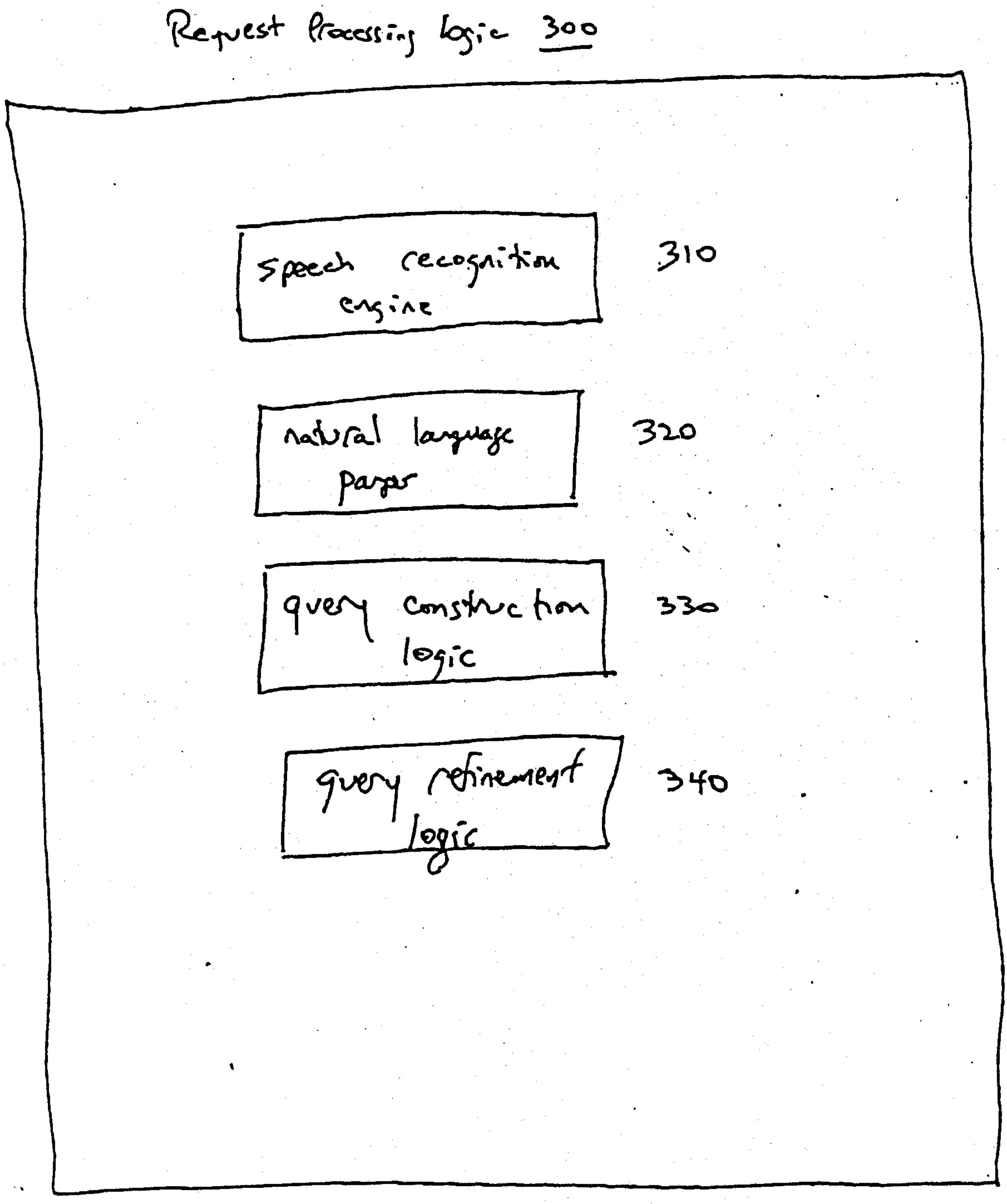

Request Processing logic 300
speech recognition engine
310
natural language parser
320
query construction logic
330
query refinement logic
340

Fig. 3

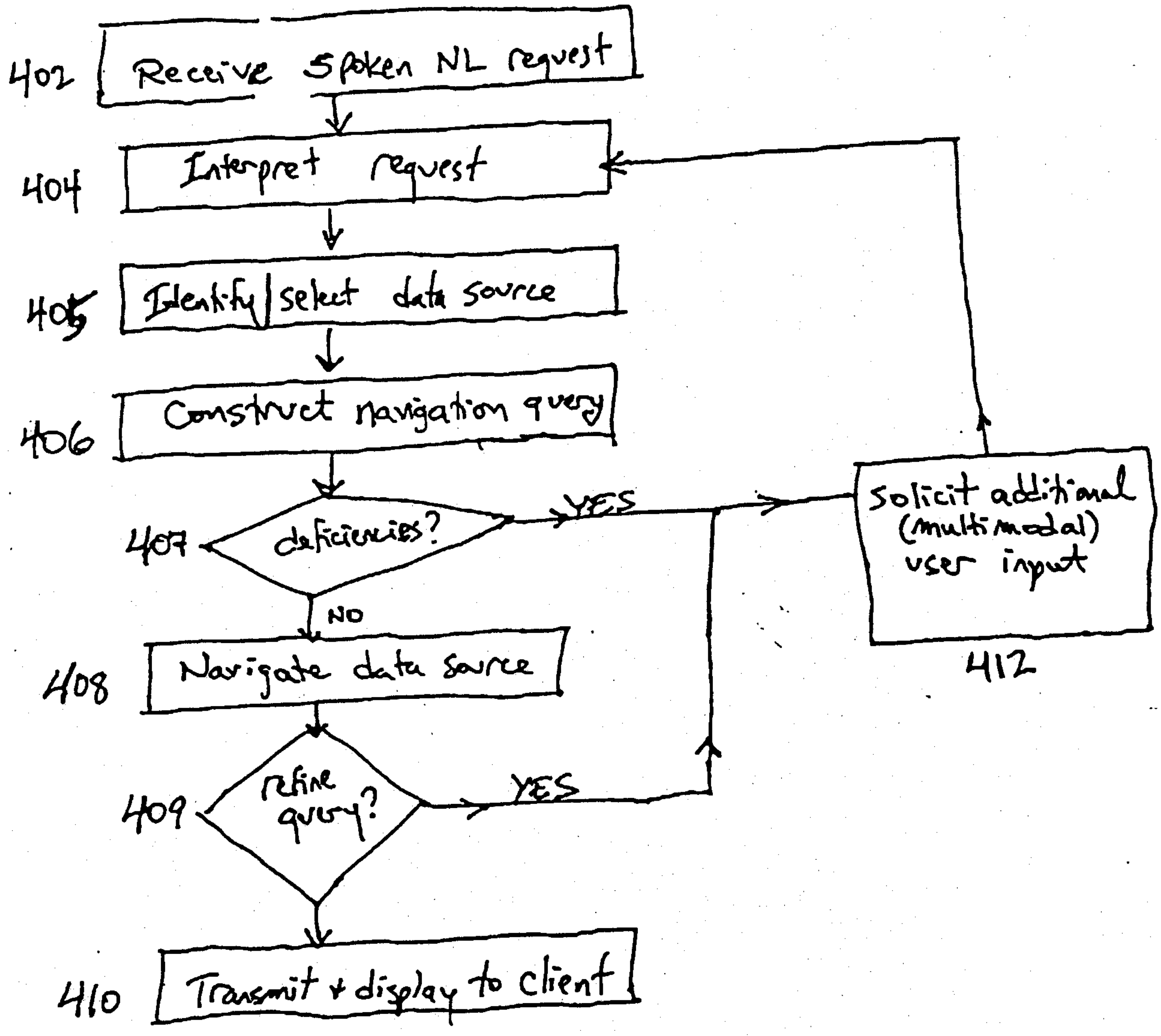

402
Receive spoken NL request
404
Interpret request
405
Identify/select data source
406
Construct navigation query
407
deficiencies?
YES
NO
Solicit additional (multimodal) user input
412
408
Navigate data source
409
refine query?
YES
410
Transmit + display to client

Fig. 4

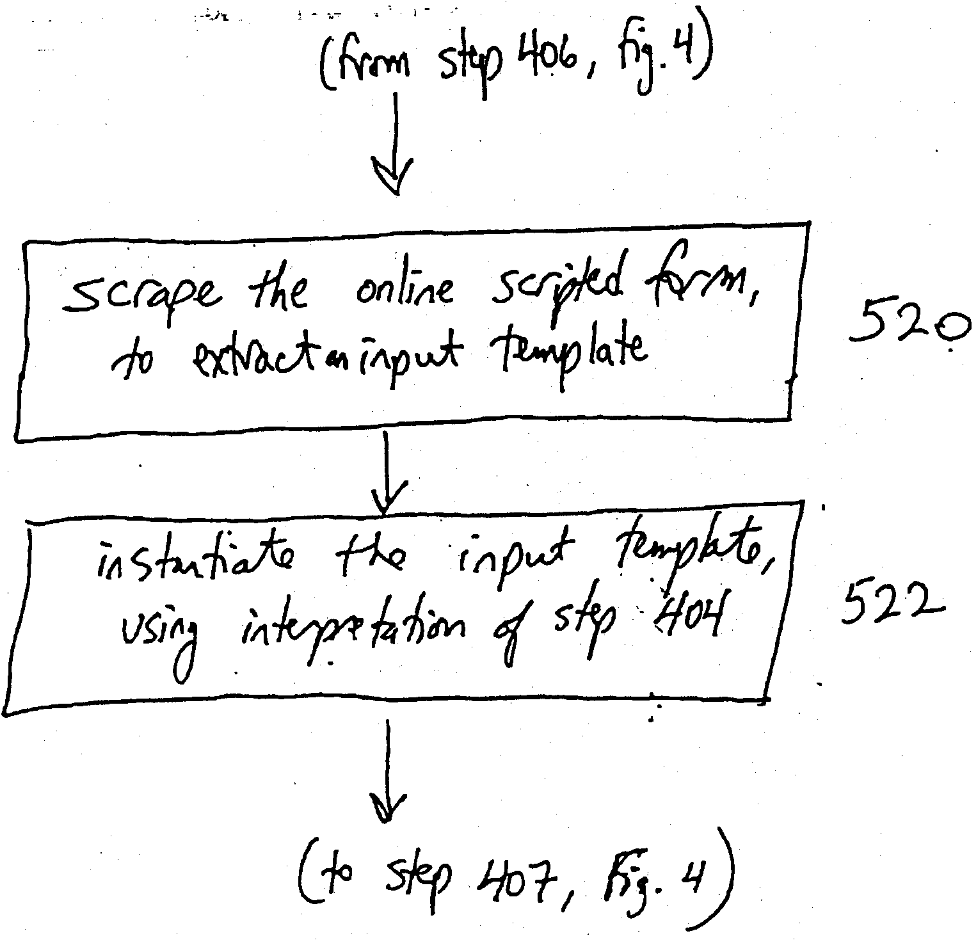

(from step 406, fig. 4)
scrape the online scripted form, to extract an input template
520
instantiate the input template, using interpretation of step 404
522
(to step 407, Fig. 4)

Fig. 5

PRINT OF DRAWINGS
AS ORIGINALLY FIL

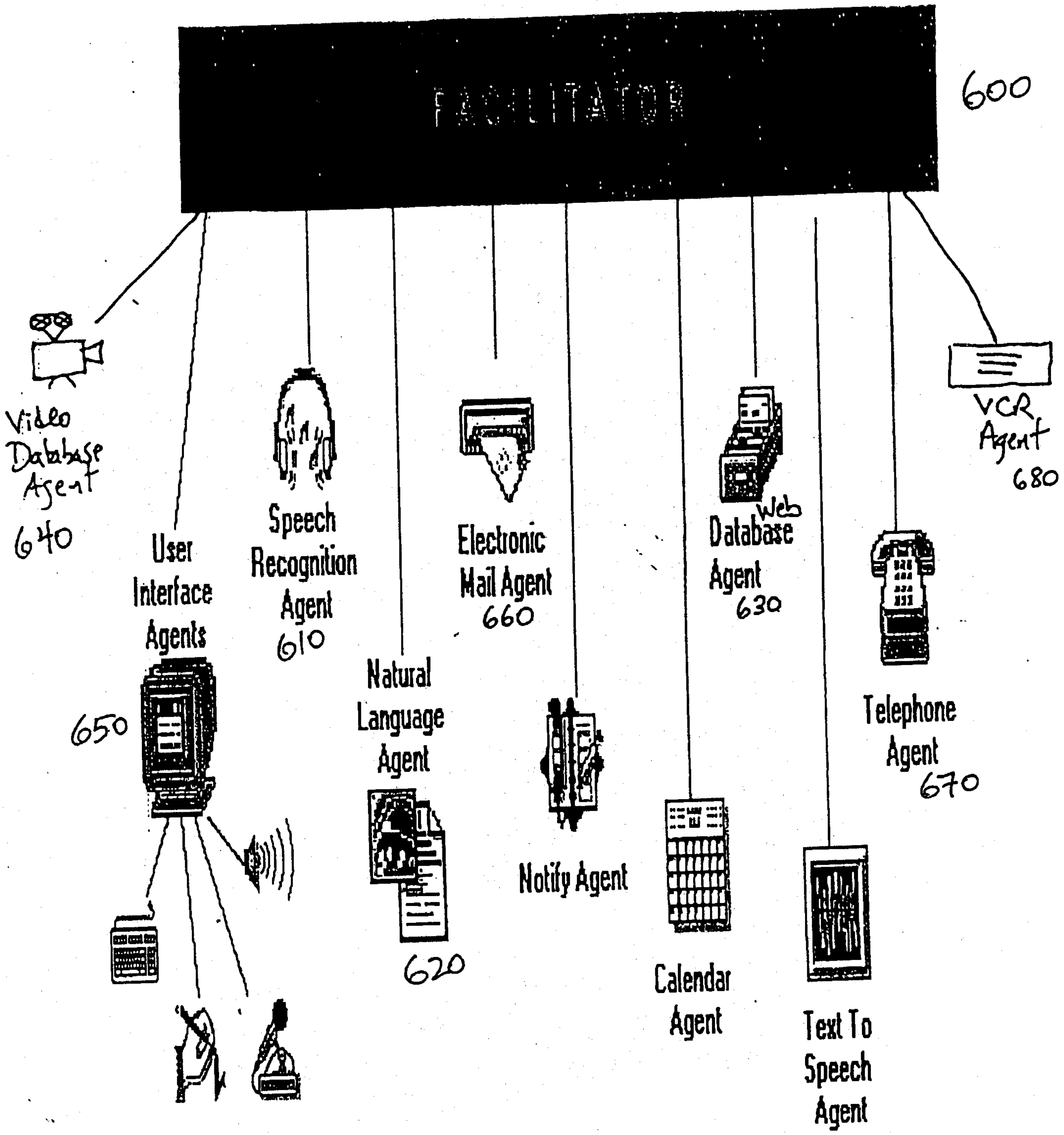


Fig. 6

**FORMALITIES LETTER**

*OC000000005113304*

**UNITED STATES DEPARTMENT OF COMMERCE**
**Patent and Trademark Office**
Address: COMMISSIONER OF PATENT AND TRADEMARKS
Washington, D.C. 20231

| APPLICATION NUMBER | FILING/RECEIPT DATE | FIRST NAMED APPLICANT | ATTORNEY DOCKET NUMBER |
|---|---|---|---|
| 09/524,095 | 03/13/2000 | Christine Halverson | SRI1P037 |

Hickman Stephens Coleman & Hughes LLP
PO Box 52037
Palo Alto, CA 94303-0746

Date Mailed: 05/12/2000

# NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

## FILED UNDER 37 CFR 1.53(b)

***Filing Date Granted***

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given TWO MONTHS from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is missing.
  *A properly signed oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date, is required.*
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of $130 for a non-small entity, must be submitted with the missing items identified in this letter.
- **The balance due by applicant is $ 130.**

---

***A copy of this notice <u>MUST</u> be returned with the reply.***

Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 3 - OFFICE COPY

04CO SE 7UR #3

**PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the application of )
) Examiner: Not Assigned
Luc Julia et al. )
) Art Unit: Not Assigned
Application No. 09/524,095 )
) Atty. Docket No. AND1P037
Filed: 3/13/2000 )
) Date: 8/17/00
For: )
Navigating Network-Based Electronic )
Information Using Spoken Natural Language )
Input With Multimodal Error Feedback )
)

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231 on August 17, 2000.

Signed: K. Main
Kimberly Main

## RESPONSE TO NOTICE TO FILE MISSING PARTS

Assistant Commissioner for Patents
**Box: Missing Parts**
Washington, D.C. 20231

Sir:

In response to the Notice to File Missing Parts of Application--Filing Date Granted dated May 12, 2000, Applicants hereby attach an original executed Declaration and Power of Attorney, an Assignment document, an Assignment Recordation Cover Sheet, and the copy of the Notice to be returned with this response. Applicants are also enclosing a copy of the previously filed Small Entity Statement, filed on the parent case of this application, serial number 09/225,198, which accounts for the fees being paid as a small entity on this case. We are also enclosing check number 6331, in the amount of $105.00, for the missing fees, and the assignment recordation. We are also request a two-month extension of time in which to responds to this matter, check number 6812, in the amount of $190.00 is also enclosed.

08/23/2000 WKOROMA 00000088 09524095
01 FC:216 190.00 OP

The Commissioner is authorized to charge any other fees that may be due to our Deposit Account No. 50-0384 (Order No. <u>SRI1P037</u>). A copy of this sheet is enclosed for this purpose.

Respectfully submitted,
HICKMAN COLEMAN & HUGHES, LLP

Raymond E. Roberts
Reg. No. 38,597

P.O. Box 52037
Palo Alto, CA 94303-0746
(408) 558-9950

# DECL ATION AND POWER OF AT RNEY
# FOR O IGINAL U.S. PATENT APPLIC TION

#3

O I P E
AUG 21 2000
JC117
PATENT & TRADEMARK OFFICE

Attorney's Docket No. SRI1P037

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: NAVIGATING NETWORK-BASED ELECTRONIC INFORMATIN USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK, the specification of which,

(check one)

1. ☐ is attached hereto.
2. ☒ was filed on March 13, 2000 as U.S. Application Serial No. 09/524,095 and was amended on ______.
3. ☐ was filed on ______ as International PCT Application Serial No. ______ and was amended on ______.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, CFR § 1.56.

I hereby claim foreign priority benefits under Title 35, United States code, § 119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed:

| **Prior Foreign Application(s)** | | | Priority Benefits Claimed? |
|---|---|---|---|
| (Appl. No.) | (Country) | (Filing Date) | ☐Yes ☐No |
| (Appl. No.) | (Country) | (Filing Date) | ☐Yes ☐No |
| (Appl. No.) | (Country) | (Filing Date) | ☐Yes ☐No |

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below:

| | |
|---|---|
| (Application Serial No.) | (Filing Date) |
| (Application Serial No.) | (Filing Date) |

I hereby claim the benefit under Title 35, United States Code, § 120 of any United States application(s), or § 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, § 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

**Prior U.S. Application(s)**

(Application Serial No.) (Filing Date) (Status - patented, pending, abandoned)

(Application Serial No.) (Filing Date) (Status - patented, pending, abandoned)

And I hereby appoint the law firm of Hickman Stephens Coleman & Hughes, including **Paul L. Hickman (Reg. No. 28,516); L. Keith Stephens (Reg. No. 32,632); Brian R. Coleman (Reg. No. 39,145); Michael J. Hughes (Reg. No. 29,077); Michael E. Melton (Reg. No. 32,276); Raymond E. Roberts (Reg. No. 38,597); Vidya R. Bhakar (Reg. No. 42,323); Larry B. Guernsey (Reg. No. 40,008); Douglas E. Mackenzie (Reg. No. 38,955); Michael D. Plimier (Reg. No. 43,004); Ronald B. Feece (Reg. No. P46,327); Stefanie M. Howell (Reg. No. P45,929); and Robert D. Hayden (Reg. No. 42,645)** as my principal attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

**Send Correspondence To:**

**HICKMAN STEPHENS COLEMAN & HUGHES, LLP**
**P.O. BOX 52037**
**Palo Alto, California 94303-0746**

**Direct Telephone Calls To:** **Raymond E. Roberts at telephone number (408) 558-9950**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Typewritten Full Name of Sole or First Inventor: Christine Halverson Citizenship: USA
**Inventor's signature:** Christine A Halverson **Date of Signature:** 6-16-00
Residence: (City) San Jose (State/Country) California/USA
Post Office Address: 1623 Fairorchard Avenue, San Jose, California 95125

Full Name of Second Joint Inventor (if any): Luc Julia Citizenship: USA
**Inventor's signature:** **Date of Signature:** 6.21.00
Residence: (City) Menlo Park (State/Country) California/USA
Post Office Address: 607 Menlo Avenue, Menlo Park, California 94025

Full Name of Third Joint Inventor (if any): Dimitris Voutsas Citizenship: Greece
**Inventor's signature:** **Date of Signature:** 6/16/00
Residence: (City) Thessaloniki (State/Country) Greece
Post Office Address: 14 M. Pyrza Street, Neoi Epivates, Thessaloniki 57019, Greece

**Prior U.S. Application(s)**

(Application Serial No.) (Filing Date) (Status - patented, pending, abandoned)

(Application Serial No.) (Filing Date) (Status - patented, pending, abandoned)

And I hereby appoint the law firm of Hickman Stephens Coleman & Hughes, including **Paul L. Hickman (Reg. No. 28,516); L. Keith Stephens (Reg. No. 32,632); Brian R. Coleman (Reg. No. 39,145); Michael J. Hughes (Reg. No. 29,077); Michael E. Melton (Reg. No. 32,276); Raymond E. Roberts (Reg. No. 38,597); Vidya R. Bhakar (Reg. No. 42,323); Larry B. Guernsey (Reg. No. 40,008); Douglas E. Mackenzie (Reg. No. 38,955); Michael D. Plimier (Reg. No. 43,004); Ronald B. Feece (Reg. No. P46,327); Stefanie M. Howell (Reg. No. P45,929); and Robert D. Hayden (Reg. No. 42,645)** as my principal attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

**Send Correspondence To:**

**HICKMAN STEPHENS COLEMAN & HUGHES, LLP**
**P.O. BOX 52037**
**Palo Alto, California 94303-0746**

**Direct Telephone Calls To:** **Raymond E. Roberts at telephone number (408) 558-9950**

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Typewritten Full Name of Sole or First Inventor: Christine Halverson
Citizenship: USA
**Inventor's signature:** Christine Halverson
**Date of Signature:** 6-16-00
Residence: (City) San Jose
(State/Country) California/USA
Post Office Address: 1623 Fairorchard Avenue, San Jose, California 95125

Full Name of Second Joint Inventor (if any): Luc Julia
Citizenship: USA
**Inventor's signature:**
**Date of Signature:**
Residence: (City) Menlo Park
(State/Country) California/USA
Post Office Address: 607 Menlo Avenue, Menlo Park, California 94025

Full Name of Third Joint Inventor (if any): Dimitris Voutsas
Citizenship: Greece
**Inventor's signature:**
**Date of Signature:** 6/16/00
Residence: (City) Thessaloniki
(State/Country) Greece
Post Office Address: 14 M. Pyrza Street, Neoi Epivates, Thessaloniki 57019, Greece

Full Name of Fourth Joint Inventor (if any): A Cheyer

Citizen USA

**Inventor's signature**: Adam J. Cheyer

**Date of Signature**: 6/22/00

Residence: (City) Palo Alto

(State/Country) California /USA

Post Office Address: 757 Cereza Drive, Palo Alto, California 94306

OIPE JC117
AUG 21 2000
PATENT & TRADEMARK OFFICE

#3

HS&C Docket No. SRI1P016
SRI Docket No. US3949-2

**PATENT**

## VERIFIED STATEMENT CLAIMING SMALL-ENTITY STATUS
## (37 CFR 1.9(f) & 1.27(d))--NONPROFIT ORGANIZATION

Applicant or Patentee: Adam J. Cheyer et al.
Serial or Patent No.: 09/225198
Filed or Issued: January 5, 1999
Title: SOFTWARE-BASED ARCHITECTURE FOR COMMUNICATION AND COOPERATION AMONG DISTRIBUTED ELECTRONIC AGENTS

I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below:

NAME OF NONPROFIT ORGANIZATION: SRI International
ADDRESS OF NONPROFIT ORGANIZATION: 333 Ravenswood Avenue
Menlo Park, CA 94025-3493

TYPE OF NONPROFIT ORGANIZATION:

☐ UNIVERSITY OR OTHER INSTITUTION OF HIGHER EDUCATION

☒ TAX-EXEMPT UNDER INTERNAL REVENUE SERVICE CODE (26 USC 501(a) and 501(c)(3))

☒ NONPROFIT SCIENTIFIC OR EDUCATIONAL UNDER STATUTE OF STATE OF THE UNITED STATES OF AMERICA
(NAME OF STATE: California)
(CITATION OF STATUTE: Sections 5110 et seq., California Corporations Code)

☐ WOULD QUALIFY AS TEXT-EXEMPT UNDER INTERNAL REVENUE SERIVCE CODE (26 USC 501(a) AND 501(c)(3)) IF LOCATED IN THE UNITED STATES OF AMERICA

☐ WOULD QUALIFY AS NONPROFIT SCIENTIFIC OR EDUCATIONAL UNDER STATUTE OF STATE OF THE UNITED STATES OF AMERICA IF LOCATED IN THE UNITED STATES OF AMERICA
(NAME OF STATE: )
(CITATION OF STATUTE: )

I hereby declare that the nonprofit organization identified above qualifies as a nonprofit organization as defined in 37 CFR 1.9(e) for purposes of paying reduced fees to the United States Patent and Trademark Office regarding the invention in:

☐ the specification filed herewith with title as listed above.

☒ the application identified above.

☐ the patent identified above.

I hereby declare that rights under contract or law have been conveyed to and remain with the non-profit organization regarding the above-identified invention. If the rights held by the nonprofit organization are not exclusive, each individual, concern, or organization having rights in the inven-tion must file separate verified statements averring to their status as small entities and that no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR

HS&C Docket No. SRI1P016
SRI Docket No. US3949-2

**PATENT**

1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:

☒ no such person, concern, or organization exists.

☐ each such person, concern, or organization is listed below.

NAME:
ADDRESS:

☐ INDIVIDUAL ☐ SMALL BUSINESS CONCERN ☐ NONPROFIT ORGANIZATION

NAME:
ADDRESS:

☐ INDIVIDUAL ☐ SMALL BUSINESS CONCERN ☐ NONPROFIT ORGANIZATION

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate (37 CFR 1.28(b)).

I hereby declare that all statements made herein of my own knowledge are true, and that all statements made on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING: Mary Lou Joyner
TITLE IN ORGANIZATION: Assistant Secretary
ADDRESS OF PERSON SIGNING: 333 Ravenswood Ave., Menlo Park, CA 94025-3493

SIGNATURE: Mary Lou Joyner

DATE: March 4, 1999

OIPE JC117
AUG 21 2000
PATENT & TRADEMARK OFFICE

#3

PATENT

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

Adam J. CHEYER et al.

Serial No. 09/225,198

Filed: January 5, 1999

For: SOFTWARE-BASED ARCHITECTURE FOR COMMUNICATION AND COOPERATION AMONG DISTRIBUTED ELECTRONIC AGENTS

Group Art Unit: 2755

Examiner: Not Assigned

Attorney Docket No. (SRI1P016)

Date: March 5, 1999

**CERTIFICATE OF MAILING**
**I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Commissioner of Patents and Trademarks, Washington, DC 20231 on March 5, 1999.**
**Signed:** Jayashree Vasudevan
**Jayashree Vasudevan**

**Commissioner of Patents and Trademarks**
**Washington, DC 20231**

ATTENTION: **Refund Section, Accounting Division, Office of Finance**

## REQUEST FOR REFUND

(Improper charge of Deposit Account)

## I. REFUND REQUEST

This is a request for a refund with respect to the charge to Deposit Account 50-0384 shown on the statement dated January 29, 1999 (Order No. SRI1P016) for the above-identified patent. A copy of the monthly statement in which the error referred to occurs, accompanies this request.

## II. FEES CHARGED FOR WHICH REFUND REQUESTED

| | |
|---|---|
| Basic Fee | $ 760.00 |
| Sixty nine (69) claims | $1242.00 |
| Three (3) Independent Claims | $ 234.00 |

for the total amount of **$2236.00** in the above referenced application.

## III. EXPLANATION OF WHY CONTESTED CHARGE IS IN ERROR

The above mentioned charges as a **large entity** were charged to our Deposit Account No. 50-0384. Enclosed herewith is a true facsimile copy of Verified Statement Claiming Small Entity Status by our client (SRI International) as a Non-Profit Organization.

## IV. MANNER OF REFUND

Please make refund by crediting Account No. 50-0384 (Order No. SRI1P016) in the amount of **$1118.00.**

Respectfully submitted,
HICKMAN STEPHENS & COLEMAN, LLP

Brian R. Coleman
Reg. No. 39,145

Hickman Stephens & Coleman, LLP
P.O. Box 52037
Palo Alto, CA 94303-0746
(650)470-7430

OIPE
AUG 2 1 2000
PATENT & TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office

Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

## ONTHLY STATEMENT F DEPOSIT ACCOUNT

replenish your Deposit Account, detach and urn top portion with your check. Make check able to Commissioner of Patents & Trademarks.

Account No. 500384
Date 1-29-99
Page 1

FINA

HICKMAN & MARTINE LLP
HUSAM Y HAMMAD
200 PAGE MILL ROAD, SUITE 100

PALO ALTO CA 94306

PLEASE SEND REMITTANCES TO:
Patent and Trademark Office
P.O. Box 70541
Chicago, Ill. 60673

| ISTED | | CONTROL NO. | DESCRIPTION (Serial, Patent, TM, Order) | DOCKET NO. | FEE CODE | CHARGES/ CREDITS | BALANCE |
|---|---|---|---|---|---|---|---|
| Y | YR. | | | | | | |
| 4 | 99 | PLH 164 | 09166819 | DHL1P001 | 103 | Reg Fee 607.00 | Refund 4417.00 |
| 5 | 99 | 52 | PCT/US98/21030 | SAS1P432.Pct | 899 | 157.50 105.00 | Call Pro 4312.00 |
| 5 | 99 | 53 | PCT/US98/21030 | | 157 | 52.50 | 4259.50 |
| 6 | 99 | 47 | 09218542 | LAM1P084 | 102 | 60.00 20.00 | 2/99 4239.50 |
| 6 | 99 | 49 | 09218542 | LAM1P084 | 581 | 40.00 | 4199.50 |
| 7 | 99 | 66 | 09220738 | ADAPP068 | 103 | 60.00 20.00 | 2/99 4179.50 |
| 7 | 99 | 68 | 09220738 | ADAPP068 | 581 | 40.00 | 4139.50 |
| 8 | 99 | 348 | 09174491 | ELECP003A | 101 | -790.00 | offset by 10/98 4929.50 |
| 8 | 99 | 351 | 09174491 | ELECP003A | 101 | 30.00 | 4899.50 Debit |
| 1 | 99 | 356 | 08766513 NOT OURS | | 704 | -55.00 | Call Pro 4954.50 |
| 2 | 99 | 38 | PAYMENT | | 701 | -5000.00 | 9954.50 |
| 2 | 99 | 144 | 60114493 | LAM1P083+ | 114 | 150.00 | 2/99 9804.50 |
| 4 | 99 | 71 | 09214694 | KKF1P004 | 704 | -90.00 | 2/99 9894.50 |
| 4 | 99 | 193 | 09173583 | ELECP006 A | 105 | 82.00 42.00 | 2/99 9852.50 |
| 14 | 99 | 195 | 09173583 | | 581 | 40.00 | 9812.50 |
| 15 | 99 | 126 | 09225198 | SRI1P016 | 581 | 40.00 | 2/99 9772.50 |
| 19 | 99 | 50 | 09225198 | SRI1P016 | 101 | 760.00 | 1/31/99 9012.50 |
| 19 | 99 | 51 | 09225198 | SRI1P016 | 102 | 234.00 | 1/31/99 8778.50 |
| 19 | 99 | 52 | 09225198 | SRI1P016 | 103 | 1242.00 | 1/31/99 7536.50 |
| 21 | 99 | 20 | 09226380 | ELECP010A | 101 | 760.00 | charged in March 6776.50 when we |
| 21 | 99 | 21 | 09226380 | ELECP010A | 102 | 234.00 | diff 6542.50 |
| 21 | 99 | 22 | 09226380 | ELECP010A | 103 | 306.00 | 6236.50 file |
| 21 | 99 | PLH 78 | PCT/US98/12578 | BEM1P001.Pct | 803 | 40.00 | 2/99 6196.50 |
| 21 | 99 | TRR 149 | 09138304 | THM1P049 | 101 | 158.00 | 2/99 6038.50 |
| 22 | 99 | BRC 64 | 09169638 | TNT1P020 | 103 | 274.00 | 2/99 5764.50 |
| 22 | 99 | PLH 70 | PCT/US98/12591 | THM1P030.Pct | 803 | 40.00 | 2/99 5724.50 |
| 22 | 99 | PLH 184 | PCT/US98/12389 | THM1P031.Pct | 803 | Call Pro 40.00 | Call Pro 5684.50 |
| 26 | 99 | 86 | 09169750 | INT1P018 | 103 | Reg. Fee 92.00 | Refund 5592.50 |
| 26 | 99 | 88 | 09169750 | | 581 | 40.00 | 5552.50 |
| 26 | 99 | 134 | 09138309 | THM1P050 | 116 | 146.00 | 2/99 5406.50 |
| 26 | 99 | TRR 136 | 09179382 | IMM1P054 | 103 | 162.00 | 2/99 5244.50 |
| 27 | 99 | 12 | 08807709 | SAS1P008B | 116 | 380.00 | 2/99 4864.50 |
| 29 | 99 | 395 | 5701140 | THM1P007.US | 145 | 100.00 | 2/99 4764.50 |

AMOUNT SUFFICIENT TO

| OPENING BALANCE | TOTAL CHARGES | TOTAL CREDITS | CLOSING BALANCE |
|---|---|---|---|
| 5024.00 | 6194.50 | 5935.00 | 4764.50 |

#3

OIPE JC117
AUG 2 1 2000
PATENT & TRADEMARK OFFICE

**PATENT POSTCARD**

**Docket No.** SRI1P016 **Appln. No.** 09/225,198 **Date:** March 5, 1999

**By:** BRC/jv **Filing Date:** January 5, 1999 **Express Mail No.:**

**Inventor(s):** Adam J. Cheyer et al.

**Title:** SOFTWARE-BASED ARCHITECTURE FOR COMMUNICATION AND COOPERATION AMONG DISTRIBUTED ELECTRONIC AGENTS

**The following has been received in the U.S. Patent & Trademark Office on the date stamped below:**

- X Return Receipt Postcard
- X Request for Refund
- X Verified Statement Claiming Small-Entity Status
- X Monthly Statement of Deposit Account dated 1/29/99

OIPE JC52
MAR 1 2 1999
PATENT & TRADEMARK OFFICE

RECD MAR 2 2 1999

## PATENT POSTCARD

Docket No. SRI1P016 Appln. No.: 09/225,198 Date: March 5, 1999

By: BRC/jv Filing Date: January 5, 1999 Express Mail No.:

Inventor(s): Adam J. Cheyer et al.

Title: SOFTWARE-BASED ARCHITECTURE FOR COMMUNICATION AND COOPERATION AMONG DISTRIBUTED ELECTRONIC AGENTS

The following has been received in the U.S. Patent & Trademark Office on the date stamped below:

X Return Receipt Postcard
X Request for Refund
X Verified Statement Claiming Small-Entity Status
X Monthly Statement of Deposit Account dated 1/29/99

OIPE JC117
AUG 2 1 2000
PATENT & TRADEMARK OFFICE

Hickman Stephens [illegible] LLP
U.S.: ✓ Foreign:
Docketed: 3/8/99 By: AP
Action: Small entity filed
Due Date: refund requested
Atty: BRC/EWJ
Docket #: SRI1P016.US

Status inquiry re refund
6/5/99

OIPE JC117
AUG 2 1 2000
PATENT & TRADEMARK OFFICE

file:///c:/APPS/preexam/correspondence/3.htm

#3

**FORMALITIES LETTER**

*OC000000005113304*

**UNITED STATES DEPARTMENT OF COMMERCE**
**Patent and Trademark Office**
Address: COMMISSIONER OF PATENT AND TRADEMARKS
Washington, D.C. 20231

| APPLICATION NUMBER | FILING/RECEIPT DATE | FIRST NAMED APPLICANT | ATTORNEY DOCKET NUMBER |
|---|---|---|---|
| 09/524,095 | 03/13/2000 | Christine Halverson | SRI1P037 |

HIckman Stephens Coleman & Hughes LLP
PO Box 52037
Palo Alto, CA 94303-0746

Date Mailed: 05/12/2000

# NOTICE TO FILE MISSING PARTS OF NONPROVISIONAL APPLICATION

## FILED UNDER 37 CFR 1.53(b)

***Filing Date Granted***

An application number and filing date have been accorded to this application. The item(s) indicated below, however, are missing. Applicant is given TWO MONTHS from the date of this Notice within which to file all required items and pay any fees required below to avoid abandonment. Extensions of time may be obtained by filing a petition accompanied by the extension fee under the provisions of 37 CFR 1.136(a).

- The oath or declaration is missing.
  *A properly signed oath or declaration in compliance with 37 CFR 1.63, identifying the application by the above Application Number and Filing Date, is required.*
- To avoid abandonment, a late filing fee or oath or declaration surcharge as set forth in 37 CFR 1.16(e) of $130 for a non-small entity, must be submitted with the missing items identified in this letter.
- **The balance due by applicant is $ 130.**

---

***A copy of this notice MUST be returned with the reply.***

Customer Service Center
Initial Patent Examination Division (703) 308-1202

PART 2 - COPY TO BE RETURNED WITH RESPONSE

08/23/2000 WKOROMA 00000088 09524095
02 FC:205 65.00 OP

#4
pre amdt
a
10/31/00
ew

**PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

OIPE JC147
JUL 0 5 2000
PATENT & TRADEMARK OFFICE

In re the application of

Christine HALVERSEN et al.

Application No. 09/524,095

Filed: March 13, 2000

For: NAVIGATING NETWORK BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK

Docket:
SRI1P037A

Date: June 30, 2000

# Preliminary Amendment

Assistant Commissioner for Patents
and Trademarks
Washington, DC 20231

Dear Sir:

In regard to the above-named patent application, please enter the following amendments.

IN THE TITLE:

Please delete "NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK", and insert therefor-- NAVIGATING NETWORK-BASED ELECTRONIC INFORMATION USING SPOKEN INPUT WITH MULTIMODAL ERROR FEEDBACK--.

IN THE ABSTRACT:

Please delete the Abstract and insert therefore --A system, method, and article of manufacture are provided for navigating an electronic data source by means of spoken language.

a¹

2

a1

When a spoken input request is received from a user, it is interpreted. Additional input is solicited from the user in a modality different than the original request and used to refine the navigation query. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources.

IN THE SPECIFICATION:

Please delete page 3, lines 3 to 32, and insert therefore, --The present invention addresses the above needs by providing a system, method, and article of manufacture for navigating network-based electronic data sources in response to spoken input requests. When a spoken input request is received from a user, it is interpreted, such as by using a speech recognition engine to extract speech data from acoustic voice signals, and using a language parser to linguistically parse the speech data. The interpretation of the spoken request can be performed on a computing device locally with the user or remotely from the user. The resulting interpretation of the request is thereupon used to automatically construct an operational navigation query to retrieve the desired information from one or more electronic network data sources, which is then transmitted to a client device of the user. If the network data source is a database, the navigation query is constructed in the format of a database query language.

a2

Typically, errors or ambiguities emerge in the interpretation of the spoken request, such that the system cannot instantiate a complete, valid navigational template. This is to be expected occasionally, and one preferred aspect of the invention is the ability to handle such errors and ambiguities in relatively graceful and user-friendly manner. Instead of simply rejecting such input and defaulting to traditional input modes or simply asking the user to try again, a preferred embodiment of the present invention seeks to converge rapidly toward instantiation of a valid navigational template by soliciting additional clarification from the user as necessary, either before or after a navigation of the data source, via multimodal input, i.e., by means of menu selection or other input modalities including and in addition to spoken input. This clarifying, multi-modal dialogue takes advantage of whatever partial navigational information has been gleaned from the initial interpretation of the user's spoken request. This clarification process continues until the system converges toward an adequately instantiated navigational template, which is in turn used to navigate the network-based data and retrieve the user's desired information. The retrieved information is transmitted across the network and presented to the user on a suitable client display device.

24

A

IN THE CLAIMS:

Please delete claims 1-55, and insert therefore the following claims 1-66:

Rule 1.126

Sub B

56 ~~1.~~ (New) A method for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken request for desired information from the user;

(b) rendering an interpretation of the spoken request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a modality different than the original request;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

R1.126

57 ~~2~~. (New) The method of claim 56 ~~1~~, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and a linguistic parser.

58 ~~3~~. (New) The method of claim 56 ~~1~~, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

59 ~~4~~. (New) The method of claim 58 ~~3~~, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

60 ~~5~~. (New) The method of claim 56 ~~1~~, wherein the navigation query is constructed in the format of a database query language.

a3

61 ~~6~~. (New) The method of claim 56 ~~1~~, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

62 ~~7~~. (New) The method of claim 56 ~~1~~, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

63 ~~8~~. (New) The method of claim 56 ~~1~~, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

64 ~~9~~. (New) The method of claim 63 ~~8~~, wherein the deficiencies include unresolved words of the spoken request.

65 ~~10~~. (New) The method of claim 63 ~~8~~, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

R1.126

66 ~~11~~. (New) The method of claim 56 ~~1~~, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

67 ~~12~~. (New) The method of claim 66 ~~11~~, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

68 ~~13~~. (New) The method of claim 66 ~~11~~, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

69 ~~14~~. (New) The method of claim 56 ~~1~~, wherein the additional input is solicited upon receiving a user-input statement that additional information is required.

70 ~~15~~. (New) The method of claim 56 ~~1~~, wherein the step of soliciting the additional input includes presenting a menu to the user on the client device of the user.

71 ~~16~~. (New) The method of claim 56 ~~1~~, wherein the step of soliciting the additional input includes presenting a textual request for the additional input.

R3

72 ~~17~~. (New) The method of claim 56 ~~1~~, wherein the step of soliciting the additional input includes an audible request for the additional input.

73 ~~18~~. (New) The method of claim 56 ~~1~~, wherein the step of soliciting the additional input includes presenting a list of portions of the electronic data source that match the navigational query.

74 ~~19~~. (New) The method of claim 56 ~~1~~, wherein additional input received from the user is at least partially speech based.

75 ~~20~~. (New) The method of claim 56 ~~1~~, wherein additional input received from the user includes no spoken input.

76 ~~21~~. (New) The method of claim 56 ~~1~~, wherein steps (d)-(e) are repeated until the navigational query is deemed adequate.

R1.126 - 77

~~22~~. (New) The method of claim ~~1~~ 56, wherein the input modality of step (d) includes selecting from a displayed option menu.

78

~~23~~. (New) The method of claim ~~22~~ 77, wherein the act of selecting from the displayed option menu is performed by speaking.

79

~~24~~. (New) The method of claim ~~1~~ 56, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

80

~~25~~. (New) The method of claim ~~1~~ 56, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

81

~~26~~. (New) The method of claim ~~1~~ 56, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

82

~~27~~. (New) A system for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

A3

(a) a portable microphone operable to receive a spoken request for desired information from the user;

(b) language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a modality different than the original request;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

R1.126

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

83 ~~28~~. (New) The system of claim 82 ~~27~~, wherein the language processing logic includes speech recognition logic and an linguistic parsing logic for deriving linguistic information.

84 ~~29~~. (New) The system of claim 82 ~~27~~, wherein the language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

85 ~~30~~. (New) The system of claim 84 ~~29~~, wherein the language processing logic dynamically scrapes the online scripted interface.

86 ~~31~~. (New) The system of claim 82 ~~27~~, wherein the query construction logic constructs the query in the format of a database query language.

a3

87 ~~32~~. (New) The system of claim 82 ~~27~~, wherein at least a portion of the language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

88 ~~33~~. (New) The system of claim 82 ~~27~~, wherein at least a portion of the language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

89 ~~34~~. (New) The system of claim 82 ~~27~~, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

90 ~~35~~. (New) The system of claim 89 ~~34~~, wherein the deficiencies include unresolved words of the spoken request.

R1.126

91 ~~36~~. (New) The system of claim ~~34~~ 82, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

92 ~~37~~. (New) The system of claim ~~27~~ 92, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

93 ~~38~~. (New) The system of claim ~~37~~ 92, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

94 ~~39~~. (New) The system of claim ~~37~~ 92, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

95 ~~40~~. (New) The system of claim ~~27~~ 82, wherein the user interaction logic displays an option menu.

a3

96 ~~41~~. (New) The system of claim ~~40~~ 95, wherein the act of selecting from the displayed option menu is performed by speaking.

97 ~~42~~. (New) The system of claim ~~27~~ 82, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

98 ~~43~~. (New) The system of claim ~~27~~ 82, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

99 ~~44~~. (New) The system of claim ~~27~~ 82, wherein the display device receives data from the electronic data source on the network servers via a communications box.

100 ~~45~~. (New) The system of claim ~~27~~ 82, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

A

R1.126

Sub C3

101 ~~46~~. (New) A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;

(e) a code segment that refines the navigation query, based upon the additional input;

a3

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

102 ~~47~~. (New) The computer program of claim 101 ~~46~~, further comprising a code segment that derives linguistic information by using a speech recognition engine and a linguistic parser.

103 ~~48~~. (New) The computer program of claim 101 ~~46~~, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

104 ~~49~~. (New) The computer program of claim 103 ~~48~~, further comprising a code segment that dynamically scrapes the online scripted interface.

105 ~~50~~. (New) The computer program of claim 101 ~~46~~, wherein the navigation query is constructed in the format of a database query language.

R1.126

106 ~~51~~. (New) The computer program of claim 101 ~~46~~, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

107 ~~52~~. (New) The computer program of claim 101 ~~46~~, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.

108 ~~53~~. (New) The computer program of claim 101 ~~46~~, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.

109 ~~54~~. (New) The computer program of claim 108 ~~53~~, wherein the deficiencies include unresolved words of the spoken request.

110 ~~55~~. (New) The computer program of claim 108 ~~53~~, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

a3

111 ~~56~~. (New) The computer program of claim 101 ~~46~~, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.

112 ~~57~~. (New) The computer program of claim 111 ~~56~~, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

113 ~~58~~. (New) The computer program of claim 112 ~~57~~, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

114 ~~59~~. (New) The computer program of claim 101 ~~46~~, wherein code segment that solicits additional input displays an option menu.

115 ~~60~~. (New) The computer program of claim 114 ~~59~~, wherein the act of selecting from the displayed option menu is performed by speaking.

R1.126

~~61~~ 116. (New) The computer program of claim ~~46~~ 101, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.

~~62~~ 117. (New) The computer program of claim ~~46~~ 101, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

~~63~~ 118. (New) The computer program of claim ~~46~~ 101, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

~~64~~ 119. (New) The computer program of claim ~~46~~ 101, wherein the additional input is solicited upon receiving a user-input statement that additional information is required.

~~65~~ 120. (New) The computer program of claim ~~46~~ 101, wherein the code segment that solicits the additional input includes a code segment that presents a menu to the user on the client device of the user.

~~66~~ 121. (New) The computer program of claim ~~46~~ 101, wherein the code segment that solicits the additional input includes a code segment that presents a textual request for the additional input.

a3

~~67~~ 122. (New) The computer program of claim ~~46~~ 101, wherein the code segment that solicits the additional input includes a code segment that produces an audible request for the additional input.

~~68~~ 123. (New) The computer program of claim ~~46~~ 101, wherein the code segment that solicits the additional input includes a code segment that presents a list of portions of the electronic data source that match the navigational query.

~~69~~ 124. (New) The computer program of claim ~~46~~ 101, wherein additional input received from the user is at least partially speech based.

~~70~~ 125. (New) The computer program of claim ~~46~~ 101, wherein additional input received from the user includes no spoken input.

a3 R1,126

126 ~~71~~. (New) The computer program of claim 101 ~~46~~, wherein code segments (d)-(e) are repeated until the navigational query is deemed adequate.

In the event a telephone conversation would expedite the prosecution of this application, the Examiner may reach the undersigned at (408) 505-5100. If any fees are due in connection with the filing of this paper, then the Commissioner is authorized to charge such fees to Deposit Account No. 50-1351 (Order No. SRI1P037A). A duplicate copy of the transmittal is enclosed for this purpose.

Respectfully submitted,

Kevin J. Zilka
Registration No. 41,429

P.O. Box 721030
San Jose, CA 95172
Telephone: (408) 505-5100

**PATENT** #8

03 Att 4

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

OIPE JC147
JUL 05 2000
PATENT & TRADEMARK OFFICE

In re the application of )
Christine HALVERSEN et al. )
Application No. 09/524,095 )
Filed: March 13, 2000 )
For: NAVIGATING NETWORK BASED ELECTRONIC INFORMATION USING SPOKEN NATURAL LANGUAGE INPUT WITH MULTIMODAL ERROR FEEDBACK )

Docket: SRI1P037A

Date: June 30, 2000

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail to: Assistant Commissioner for Patents, Washington, DC 20231 on June 30, 2000.

Signed: ______________________
Kevin J. Zilka

Assistant Commissioner for Patents
Box Fee Amendment
Washington, DC 20231

Sir:

Transmitted herewith is an amendment in the above-identified application.

The fee has been calculated as shown below.

| | Claims Remaining After Amendment | | Highest Previously Paid For Extra | Present | SMALL ENTITY RATE | FEE | | LARGE ENTITY RATE | FEE |
|---|---|---|---|---|---|---|---|---|---|
| TOTAL CLAIMS | 71 | - | 55 | 16 | X09 = | $ 144 | OR | X18 = | $ |
| INDEP CLAIMS | 3 | - | 3 | 0 | X39 = | $ | OR | X78 = | $ |
| [ ] Multiple Dependent Claim Present and Fee Not Previously Paid | | | | | | $130 | | | $260 |
| | | | | TOTAL | | $ $144.00 | | | $ |

☐ Applicant(s) hereby petition for a month extension of time to respond to the outstanding Office Action.

☐ Applicant(s) believe that no (additional) Extension of Time is required; however, if it is determined that such an extension is required, Applicant(s) hereby petition that such an extension be granted and authorize the Commissioner to charge the required fees for an Extension of Time under 37 CFR 1.136 to Deposit Account No. 50-1351.

☒ Enclosed is our Check No. 139 in the amount of $144.00 to cover the additional claim fee and/or extension of time fees.

☒ If the required fees are missing or any additional fees are required to facilitate filing the enclosed response, please charge such fees or credit any overpayment to Deposit Account No. 50-1351 (Order No. SRI1P037A).

Respectfully submitted,

Kevin J. Zilka
Registration No. 41,429

P.O. Box 721030
San Jose, CA 95172
Telephone: (408) 505-5100

(Revised 1/96)

OIPE JC38
JUL 26 2000
PATENT & TRADEMARK OFFICE

SECTOR #5

**PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: )
Christine Halverson et al.. )
) Group Art Unit: Unknown
)
Application No. 09/524,095 ) Examiner: Unknown
)
Filed: 3/13/00 ) Date: July 17, 2000
)
For: Navigating Network-Based Electronic Information )
Using Spoken Natural Language Input With Multimodal )
Error Feedback )
)

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents and Trademarks, Washington, DC 20231 on July 17, 2000.

Signed: ______________________
Kimberly Main

## REQUEST FOR STATUS

Assistant Commissioner for Patents
Washington, D. C. 20231

Sir:

Applicant hereby requests status of the above-referenced patent application. This application was filed on March 13, 2000, and no Notice of Missing parts has been received as of this date.

Respectfully submitted

HICKMAN STEPHENS COLEMAN & HUGHES, LLP

Raymond E. Roberts
Reg. No. 38,597

P.O. Box 52037
Palo Alto, CA 94303-0746
(408) 558-9950

Attorney Docket No. SRI1P037

GP 2758

**PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

w/refs
#6 IDS
10/31/00
dw

OIPE JC103
MAY 2 6 2000
PATENT & TRADEMARK OFFICE

RECEIVED
JUN - 1 2000
TC 2700 MAIL ROOM

In re the application of: )
)
Christine Halverson )
)
Application No.: Unassigned 09524095 )
)
Filed: 3/13/2000 )
)
For: Navigating Network-Based Electronic Information Using Spoken Natural Language Input with Multimodal Error Feedback )

Group Art Unit: Unknown

Examiner: Unknown

Atty. Docket No.: SRI1P03

Date: May 23, 2000

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, DC 20231 on May 23, 2000.

Signed: Kimberly Main

## INFORMATION DISCLOSURE STATEMENT
## UNDER 37 CFR §§ 1.56 AND 1.97(c)

Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

The references listed in the attached PTO Form 1449, copies of which are attached, may be material to examination of the above-identified patent application. Applicants submit these references in compliance with their duty of disclosure pursuant to 37 CFR §§ 1.56 and 1.97. The Examiner is requested to make these references of official record in this application.

This Information Disclosure Statement is not to be construed as a representation that a search has been made, that additional information material to the examination of this application does not exist, or that these references indeed constitute prior art.

It is believed that no fees are due in connection with the filing of this Information Disclosure Statement. However, if it is determined that any fees are due, the Commissioner is hereby authorized to charge such fees to Deposit Account 50-0384 (Order No. SRI1P037).

Respectfully submitted,
HICKMAN STEPHENS COLEMAN & HUGHES, LLP

L. Keith Stephens
Reg. No. 32,632

P.O. Box 52037
Palo Alto, CA 94303-0746
Telephone: (408) 558-9950

| | | |
|---|---|---|
| **Form 1449 (Modified)**<br>**Information Disclosure Statement By Applicant**<br>(Use Several Sheets if Necessary) | Atty. Docket No.<br>SRI1P037<br>Applicant:<br>Christine Halverson<br>Filing Date:<br>3/13/2000 | Application No.:<br>~~Unassigned~~ 09524095<br>Group Art Unit:<br>Unknown |

### U.S. Patent Documents

| Examiner Initial | No. | Patent No. | Date | Patentee | Class | Sub-class | Filing Date |
|---|---|---|---|---|---|---|---|
| F.B. | A | 5,197,005 | 3/23/93 | Schwartz et al. | 364 | 419 | |
| | B | 5,386,556 | 1/31/95 | Hedin et al. | 395 | 600 | |
| | C | 5,434,777 | 7/18/95 | Luciw | 364 | 419 | |
| | D | 5,519,608 | 5/21/96 | Kupiec | 364 | 419.08 | |
| | E | 5,608,624 | 3/4/97 | Luciw | 395 | 794 | |
| | F | 5,721,938 | 2/24/98 | Stuckey | 395 | 754 | |
| | G | 5,729,659 | 3/17/98 | Potter | 395 | 2.79 | |
| | H | 5,748,974 | 5/5/98 | Johnson | 395 | 759 | |
| | I | 5,774,859 | 6/30/98 | Houser et al. | 704 | 275 | |
| | J | 5,794,050 | 8/11/98 | Dahlgren et al. | 395 | 708 | |
| | K | 5,802,526 | 9/1/98 | Fawcett et al. | 707 | 104 | |

### Foreign Patent or Published Foreign Patent Application

| Examiner Initial | No. | Document No. | Publication Date | Country or Patent Office | Class | Sub-class | Translation | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Yes | No |
| | L | | | | | | | |
| | M | | | | | | | |
| | N | | | | | | | |
| | O | | | | | | | |
| | P | | | | | | | |

### Other Documents

| Examiner Initial | No. | Author, Title, Date, Place (e.g. Journal) of Publication |
|---|---|---|
| | R | |
| | S | |
| | T | |

| | |
|---|---|
| Examiner | Date Considered<br>4/6/01 |

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

| | |
|---|---|
| **Form 1449 (Modified)**<br>**Information Disclosure Statement By Applicant**<br>(Use Several Sheets if Necessary) | Atty. Docket No.<br>SRI1P037<br>Applicant:<br>Christine Halverson<br>Filing Date:<br>3/13/2000 |
| | Application No.:<br>~~Unassigned~~ 09/523,945<br>Group Art Unit:<br>Unknown |

### U.S. Patent Documents

| Examiner Initial | No. | Patent No. | Date | Patentee | Class | Sub-class | Filing Date |
|---|---|---|---|---|---|---|---|
| F.B. | A | 5,805,775 | 9/8/98 | Eberman et al. | 395 | 12 | |
| | B | 5,855,002 | 12/29/98 | Armstrong | 704 | 270 | |
| | C | 5,890,123 | 3/30/99 | Brown et al. | 704 | 275 | |
| | D | 5,963,940 | 10/5/99 | Liddy et al. | 707 | 5 | |
| | E | 6,003,072 | 12/14/99 | Gerritsen et al. | 709 | 218 | |
| | F | 6,012,030 | 1/4/00 | French-St. George et al. | 704 | 275 | |
| | G | 6,026,388 | 2/15/00 | Liddy et al. | 707 | 1 | |
| | H | | | | | | |
| | I | | | | | | |
| | J | | | | | | |
| | K | | | | | | |

### Foreign Patent or Published Foreign Patent Application

| Examiner Initial | No. | Document No. | Publication Date | Country or Patent Office | Class | Sub-class | Translation | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Yes | No |
| | L | | | | | | | |
| | M | | | | | | | |
| | N | | | | | | | |
| | O | | | | | | | |
| | P | | | | | | | |

### Other Documents

| Examiner Initial | No. | Author, Title, Date, Place (e.g. Journal) of Publication |
|---|---|---|
| | R | |
| | S | |

| | |
|---|---|
| Examiner Firmin Backer | Date Considered 4/6/01 |

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

**PATENT**

jc530 U.S. PTO
09/524095
03/13/00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the application of: )
)
Christine Halverson )
)
Application No.: Unassigned )
)
Filed: Herewith )
)
For: Navigating Network-Based Electronic Information Using Spoken Natural Language Input with Multimodal Error Feedback )

Group Art Unit: Unknown

Examiner: Unknown

Atty. Docket No.: SRI1P037

Date: March 13, 2000

#7 IDS w/refs
10/31/00
JW

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, DC 20231 on March 13, 2000.

Signed: ______________________
Julie A. Curts

## INFORMATION DISCLOSURE STATEMENT
## UNDER 37 CFR §§ 1.56 AND 1.97(c)

Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

The references listed in the attached PTO Form 1449, copies of which are attached, may be material to examination of the above-identified patent application. Applicants submit these references in compliance with their duty of disclosure pursuant to 37 CFR §§ 1.56 and 1.97. The Examiner is requested to make these references of official record in this application.

This Information Disclosure Statement is not to be construed as a representation that a search has been made, that additional information material to the examination of this application does not exist, or that these references indeed constitute prior art.

It is believed that no fees are due in connection with the filing of this Information Disclosure Statement. However, if it is determined that any fees are due, the Commissioner is hereby authorized to charge such fees to Deposit Account 50-0384 (Order No. SRI1P037).

Respectfully submitted,
HICKMAN STEPHENS COLEMAN & HUGHES, LLP

Dominic M. Kotab
Reg. No. 42,762

P.O. Box 52037
Palo Alto, CA 94303-0746
Telephone: (408) 558-9950

THE FOLLOWING WAS MISSING

FROM THE ORIGINAL USPTO

FILE HISTORY

PTO-1449

Pages **1 of 3** & **3 of 3**

| | | |
|---|---|---|
| **Form 1449 (Modified)**<br>**Information Disclosure Statement By Applicant**<br>(Use Several Sheets if Necessary) | Atty. Docket No.<br>SRI1P037<br>Applicant:<br>Christine Halverson<br>Filing Date:<br>Herewith | Application No.:<br>~~Unassigned~~<br>09/824,665<br><br>Group Art Unit:<br>Unknown |

### U.S. Patent Documents

| Examiner Initial | No. | Patent No. | Date | Patentee | Class | Sub-class | Filing Date |
|---|---|---|---|---|---|---|---|
| | A | | | | | | |
| | B | | | | | | |
| | C | | | | | | |
| | D | | | | | | |
| | E | | | | | | |
| | F | | | | | | |
| | G | | | | | | |
| | H | | | | | | |
| | I | | | | | | |
| | J | | | | | | |
| | K | | | | | | |

### Foreign Patent or Published Foreign Patent Application

| Examiner Initial | No. | Document No. | Publication Date | Country or Patent Office | Class | Sub-class | Translation | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Yes | No |
| | L | | | | | | | |
| | M | | | | | | | |
| | N | | | | | | | |
| | O | | | | | | | |
| | P | | | | | | | |

### Other Documents

| Examiner Initial | No. | Author, Title, Date, Place (e.g. Journal) of Publication |
|---|---|---|
| F.B | R | http://www.ai.sri.com/~lesaf/commandtalk.html: "CommandTalk: A Spoken-Language Interface for Battlefield Simulations", 1997, by Robert Moore, John Dowding, Harry Bratt, J. Mark Gawron, Yonael Gorfu and Adam Cheyer, in "Proceedings of the Fifth Conference on Applied Natural Language Processing", Washington, DC, pp. 1-7, Association for Computational Linguistics |
| F.B | S | "The CommandTalk Spoken Dialogue System", 1999, by Amanda Stent, John Dowding, Jean Mark Gawron, Elizabeth Owen Bratt and Robert Moore, in "Proceedings of the Thirty-Seventh Annual Meeting of the ACL", pp. 183-190, University of Maryland, College Park, MD, Association for Computational Linguistics |

| | |
|---|---|
| Examiner | Date Considered<br>12/30/02 |

Examiner: Initial citation considered. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

OIPE
SEP 20 2000
PATENT & TRADEMARK OFFICE

#7/a/

PATENT B

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

2031 55

RECEIVED
APR 12 2001
Technology Center 2100

In re the application of )
)
Christine HALVERSEN et al. )
)
Application No. 09/524,095 )
)
Filed: March 13, 2000 )
)
For: NAVIGATING NETWORK BASED ELECTRONIC INFORMATION USING SPOKEN INPUT WITH MULTIMODAL ERROR FEEDBACK )

Docket:
SRI1P037A

Date: September 12, 2000

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, DC 20231 on September 12, 2000.

Signed: ____________________
Kevin J. Zilka

09/22/2000 EFLORES 00000035 09524095

01 FC:203 495.00 OP
02 FC:202 117.00 OP

## Preliminary Amendment B

Assistant Commissioner for Patents
and Trademarks
Washington, DC 20231

Dear Sir:

Please supplement the Preliminary Amendment filed June 30, 2000 regarding the above-identified patent application by entering the following amendments.

IN THE CLAIMS:

Please re-insert the originally filed claims as new claims 72-126. Pending claims 1-71 added in the previous Preliminary Amendment have been included for reference purposes. All currently pending claims are thus represented below.

1. A method for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

(a) receiving a spoken request for desired information from the user;

(b) rendering an interpretation of the spoken request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a modality different than the original request;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

2. The method of claim 1, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and a linguistic parser.

3. The method of claim 1, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

4. The method of claim 3, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

5. The method of claim 1, wherein the navigation query is constructed in the format of a database query language.

6. The method of claim 1, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

7. The method of claim 1, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

8. The method of claim 1, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

9. The method of claim 8, wherein the deficiencies include unresolved words of the spoken request.

10. The method of claim 8, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

11. The method of claim 1, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

12. The method of claim 11, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

13. The method of claim 11, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

14. The method of claim 1, wherein the additional input is solicited upon receiving a user-input statement that additional information is required.

15. The method of claim 1, wherein the step of soliciting the additional input includes presenting a menu to the user on the client device of the user.

16. The method of claim 1, wherein the step of soliciting the additional input includes presenting a textual request for the additional input.

17. The method of claim 1, wherein the step of soliciting the additional input includes an audible request for the additional input.

18. The method of claim 1, wherein the step of soliciting the additional input includes presenting a list of portions of the electronic data source that match the navigational query.

19. The method of claim 1, wherein additional input received from the user is at least partially speech based.

20. The method of claim 1, wherein additional input received from the user includes no spoken input.

21. The method of claim 1, wherein steps (d)-(e) are repeated until the navigational query is deemed adequate.

22. The method of claim 1, wherein the input modality of step (d) includes selecting from a displayed option menu.

23. The method of claim 22, wherein the act of selecting from the displayed option menu is performed by speaking.

24. The method of claim 1, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

25. The method of claim 1, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

26. The method of claim 1, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

27. A system for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken request for desired information from the user;

(b) language processing logic, operable to render an interpretation of the spoken request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a modality different than the original request;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

28. The system of claim 27, wherein the language processing logic includes speech recognition logic and an linguistic parsing logic for deriving linguistic information.

29. The system of claim 27, wherein the language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

30. The system of claim 29, wherein the language processing logic dynamically scrapes the online scripted interface.

31. The system of claim 27, wherein the query construction logic constructs the query in the format of a database query language.

32. The system of claim 27, wherein at least a portion of the language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

33. The system of claim 27, wherein at least a portion of the language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

34. The system of claim 27, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

35. The system of claim 34, wherein the deficiencies include unresolved words of the spoken request.

36. The system of claim 34, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.

37. The system of claim 27, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

38. The system of claim 37, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

39. The system of claim 37, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

40. The system of claim 27, wherein the user interaction logic displays an option menu.

41. The system of claim 40, wherein the act of selecting from the displayed option menu is performed by speaking.

42. The system of claim 27, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.

43. The system of claim 27, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

44. The system of claim 27, wherein the display device receives data from the electronic data source on the network servers via a communications box.

45. The system of claim 27, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

46. A computer program embodied on a computer readable medium for speech-based navigation of an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken request;

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;

(e) a code segment that refines the navigation query, based upon the additional input;

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

47. The computer program of claim 46, further comprising a code segment that derives linguistic information by using a speech recognition engine and a linguistic parser.

48. The computer program of claim 46, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

49. The computer program of claim 48, further comprising a code segment that dynamically scrapes the online scripted interface.

50. The computer program of claim 46, wherein the navigation query is constructed in the format of a database query language.

51. The computer program of claim 46, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

52. The computer program of claim 46, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.

53. The computer program of claim 46, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.

54. The computer program of claim 53, wherein the deficiencies include unresolved words of the spoken request.

~~55. The computer program of claim 53, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken request.~~

~~56. The computer program of claim 46, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.~~

~~57. The computer program of claim 56, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.~~

~~58. The computer program of claim 57, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.~~

~~59. The computer program of claim 46, wherein code segment that solicits additional input displays an option menu.~~

~~60. The computer program of claim 59, wherein the act of selecting from the displayed option menu is performed by speaking.~~

~~61. The computer program of claim 46, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.~~

~~62. The computer program of claim 46, further comprising a code segment that selects the data source, from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken request.~~

~~63. The computer program of claim 46, wherein the electronic data source stores multimedia content including at least one of video content and audio content.~~

~~64. The computer program of claim 46, wherein the additional input is solicited upon receiving a user-input statement that additional information is required.~~

65. The computer program of claim 46, wherein the code segment that solicits the additional input includes a code segment that presents a menu to the user on the client device of the user.

66. The computer program of claim 46, wherein the code segment that solicits the additional input includes a code segment that presents a textual request for the additional input.

67. The computer program of claim 46, wherein the code segment that solicits the additional input includes a code segment that produces an audible request for the additional input.

68. The computer program of claim 46, wherein the code segment that solicits the additional input includes a code segment that presents a list of portions of the electronic data source that match the navigational query.

69. The computer program of claim 46, wherein additional input received from the user is at least partially speech based.

70. The computer program of claim 46, wherein additional input received from the user includes no spoken input.

71. The computer program of claim 46, wherein code segments (d)-(e) are repeated until the navigational query is deemed adequate.

---

127 ~~72~~. (New) A method for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising the steps of:

B1

(a) receiving a spoken natural language ("NL") request for desired information from the user;

(b) rendering an interpretation of the spoken natural language request;

(c) constructing at least part of a navigation query based upon the interpretation;

(d) soliciting additional input from the user, including user interaction in a modality different than the original request;

(e) refining the navigation query, based upon the additional input;

(f) using the refined navigation query to select a portion of the electronic data source; and

(g) transmitting the selected portion of the electronic data source from the network server to a client device of the user.

128 ~~73.~~ (New) The method of claim ~~72~~ 127, wherein the step of rendering an interpretation further includes deriving linguistic information by using a speech recognition engine and an NL parser.

129 ~~74.~~ (New) The method of claim ~~72~~ 127, wherein the step of constructing a navigation query further includes the steps of extracting an input template for an online scripted interface to the data source, and using the input template to construct the navigation query.

B1

130 ~~75.~~ (New) The method of claim ~~74~~ 129, wherein the step of extracting an input template includes dynamically scraping the online scripted interface.

131 ~~76.~~ (New) The method of claim ~~72~~ 127, wherein the navigation query is constructed in the format of a database query language.

132 ~~77.~~ (New) The method of claim ~~72~~ 127, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a computing device located locally with the user.

133 ~~78.~~ (New) The method of claim ~~72~~ 127, wherein the step of rendering an interpretation and the step of constructing a navigation query are performed, at least in part, on a network computing device located remotely from the user.

134 ~~79.~~ (New) The method of claim ~~72~~ 127, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered during the step of constructing a navigation query.

135 ~~80~~. (New) The method of claim 134 ~~79~~, wherein the deficiencies include unresolved words of the spoken NL request.

136 ~~81~~. (New) The method of claim 134 ~~79~~, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

137 ~~82~~. (New) The method of claim 127 ~~72~~, wherein the step of soliciting additional input is performed in response to one or more deficiencies encountered after a first navigation of the data source using the navigation query constructed in step (c).

138 ~~83~~. (New) The method of claim ~~82~~, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

139 ~~84~~. (New) The method of claim 137 ~~82~~, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

B1

140 ~~85~~. (New) The method of claim 127 ~~72~~, wherein the input modality of step (d) includes selecting from a displayed option menu.

141 ~~86~~. (New) The method of claim 140 ~~85~~, wherein the act of selecting from the displayed option menu is performed by speaking.

142 ~~87~~. (New) The method of claim 127 ~~72~~, wherein the method is performed with respect to a plurality of simultaneous users and corresponding client devices.

143 ~~88~~. (New) The method of claim 127 ~~72~~, further including the step of selecting the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

144 ~~89~~. (New) The method of claim 127 ~~72~~, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

145 ~~90~~. (New) A system for utilizing spoken natural language to navigate an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, the system comprising:

(a) a portable microphone operable to receive a spoken natural language ("NL") request for desired information from the user;

(b) spoken language processing logic, operable to render an interpretation of the spoken natural language request;

(c) query construction logic, operable to construct a navigation query in response to the interpretation of the spoken natural language request;

(d) user interaction logic, operable to solicit additional input from the user, including user interaction in a modality different than the original request;

(e) query refining logic, operable to refine the navigation query, based upon the additional input;

(f) navigation logic, operable to select a portion of the electronic data source using the navigation query; and

B1

(g) electronic communications infrastructure for transmitting the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

146 ~~91~~. (New) The system of claim 145 ~~90~~, wherein the spoken language processing logic includes speech recognition logic and an NL parsing logic for deriving linguistic information.

147 ~~92~~. (New) The system of claim 145 ~~90~~, wherein the spoken language processing logic extracts an input template for an online scripted interface to the data source, and uses the input template to construct the navigation query.

148 ~~93~~. (New) The system of claim 147 ~~92~~, wherein the spoken language processing logic dynamically scrapes the online scripted interface.

149 ~~94~~. (New) The system of claim 145 ~~90~~, wherein the query construction logic constructs the query in the format of a database query language.

150 ~~95~~. (New) The system of claim 145 ~~90~~, wherein at least a portion of the spoken language processing logic is hosted on a computing device located locally with the user, and wherein the portable microphone is electronically coupled to the local computing device.

151 ~~96~~. (New) The system of claim 145 ~~90~~, wherein at least a portion of the spoken language processing logic is hosted on a network computing device located remotely from the user, and wherein the portable microphone sends data to the remote network computing device via the communications infrastructure.

152 ~~97~~. (New) The system of claim 145 ~~90~~, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered during construction of the navigation query.

153 ~~98~~. (New) The system of claim 97, wherein the deficiencies include unresolved words of the spoken NL request.

B'

154 ~~99~~. (New) The system of claim 97, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

155 ~~100~~. (New) The system of claim 145 ~~90~~, wherein the user interaction logic solicits additional input in response to one or more deficiencies encountered after a first navigation of the data source performed by the navigation logic.

156 ~~101~~. (New) The system of claim ~~100~~, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

157 ~~102~~. (New) The system of claim ~~100~~, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

158 ~~103~~. (New) The system of claim 145 ~~90~~, wherein the user interaction logic displays an option menu.

159 ~~104~~. (New) The system of claim ~~103~~, wherein the act of selecting from the displayed option menu is performed by speaking.

160 ~~105~~. (New) The system of claim 145 ~~90~~, wherein the navigation logic selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

161 ~~106~~. (New) The system of claim 145 ~~90~~, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

162 ~~107~~. (New) The system of claim 145 ~~90~~, wherein the display device receives data from the electronic data source on the network servers via a communications box.

163 ~~108~~. (New) The system of claim 145 ~~90~~, wherein the electronic communication infrastructure is a two-way infrastructure and is selected from among one or more of the following group: {coaxial cable, DSL, satellite, wireless/cellular, fiber-optic}.

B1

164 ~~109~~. (New) A computer program embodied on a computer readable medium for utilizing spoken natural language for navigating an electronic data source, the electronic data source being located at one or more network servers located remotely from a user, comprising:

(a) a code segment that receives a spoken natural language ("NL") request for desired information from the user;

(b) a code segment that renders an interpretation of the spoken natural language request;

(c) a code segment that constructs at least part of a navigation query based upon the interpretation;

(d) a code segment that solicits additional input from the user, including user interaction in a modality different than the original request;

(e) a code segment that refines the navigation query, based upon the additional input;

(f) a code segment that uses the refined navigation query to select a portion of the electronic data source; and

(g) a code segment that transmits the selected portion of the electronic data source from the network server to a primarily stationary, display device located locally with the user.

165 ~~110~~. (New) The computer program of claim 164 ~~109~~, further comprising a code segment that derives linguistic information by using a speech recognition engine and an NL parser.

166 ~~111~~. (New) The computer program of claim 164 ~~109~~, further comprising a code segment that extract an input template for an online scripted interface to the data source, and a code segment that uses the input template to construct the navigation query.

B1

167 ~~112~~. (New) The computer program of claim 166 ~~111~~, further comprising a code segment that dynamically scrapes the online scripted interface.

168 ~~113~~. (New) The computer program of claim 164 ~~109~~, wherein the navigation query is constructed in the format of a database query language.

169 ~~114~~. (New) The computer program of claim 164 ~~109~~, wherein rendering of the interpretation and the construction of the navigation query are performed, at least in part, on a computing device located locally with the user.

170 ~~115~~. (New) The computer program of claim 164 ~~109~~, wherein the rendering of the interpretation and the construction of a navigation query are performed, at least in part, on a network computing device located remotely from the user.

171 ~~116~~. (New) The computer program of claim 164 ~~109~~, wherein code segment that solicits additional input solicits the additional input in response to one or more deficiencies encountered during the constructing of the navigation query.

172 ~~117~~. (New) The computer program of claim 171 ~~116~~, wherein the deficiencies include unresolved words of the spoken NL request.

173 ~~118~~. (New) The computer program of claim 171 ~~116~~, wherein the deficiencies include one or more required elements of the navigational query not determinable from the interpretation of the spoken NL request.

174 ~~119~~. (New) The computer program of claim 164 ~~109~~, wherein the code segment that solicits the additional input solicits the additional input in response to one or more deficiencies encountered after a first navigation of the data source.

175 ~~120~~. (New) The computer program of claim 174 ~~119~~, wherein the deficiencies include existence of more than one data record within the data source responsive to the navigation query.

176 ~~121~~. (New) The computer program of claim 174 ~~119~~, wherein the deficiencies include failure to identify a single data record within the data source responsive to the navigation query.

B1

177 ~~122~~. (New) The computer program of claim 164 ~~109~~, wherein code segment that solicits additional input displays an option menu.

178 ~~123~~. (New) The computer program of claim 177 ~~122~~, wherein the act of selecting from the displayed option menu is performed by speaking.

179 ~~124~~. (New) The computer program of claim 164 ~~109~~, wherein the code segments of the computer program operate with respect to a plurality of simultaneous users and corresponding client devices.

180 ~~125~~. (New) The computer program of claim 164 ~~109~~, further comprising a code segment that selects the data source from among a plurality of candidate electronic data sources, in response to the interpretation of the spoken NL request.

181 ~~126~~. (New) The computer program of claim 164 ~~109~~, wherein the electronic data source stores multimedia content including at least one of video content and audio content.

---

REMARKS

In the event a telephone conversation would expedite the prosecution of this application, the Examiner may reach the undersigned at (408) 505-5100. If any fees are due in connection with the filing of this paper, then the Commissioner is authorized to charge such fees to Deposit Account No. 50-1351 (Order No. SRI1P037A).

Respectfully submitted,

Kevin J. Zilka
Registration No. 41,429
SILICON VALLEY IP LAW GROUP

P.O. Box 721030
San Jose, CA 95172
Telephone: (408) 505-5100

~~PATENT~~ B4

2155

RECEIVED

APR 1 2 2001

Technology Center 2100

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

OIPE JC147

SEP 2 0 2000

PATENT & TRADEMARK OFFICE

In re the application of )

Christine HALVERSEN et al. )

Application No. 09/524,095 )

Filed: March 13, 2000 )

For: NAVIGATING NETWORK BASED ELECTRONIC INFORMATION USING SPOKEN INPUT WITH MULTIMODAL ERROR FEEDBACK )

Docket:
SRI1P037A

Date: September 12, 2000

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail to: Assistant Commissioner for Patents, Washington, DC 20231 on September 12, 2000.

Signed: ________________
Kevin J. Zilka

Match & Return

Assistant Commissioner for Patents
Box Fee Amendment
Washington, DC 20231

Sir:

Transmitted herewith is an amendment in the above-identified application.

The fee has been calculated as shown below.

| | Claims Remaining After Amendment | | Highest Previously Paid For | Present Extra | SMALL ENTITY RATE FEE | OR | LARGE ENTITY RATE FEE |
|---|---|---|---|---|---|---|---|
| TOTAL CLAIMS | 126 | - | 71 | 55 | X09 = $ 495.00 OR | X18 = | $ = 203 x 55<br>202 x 3 |
| INDEP CLAIMS | 6 | - | 3 | 3 | X39 = $ 117.00 OR | X78 = | $ |
| [ ] Multiple Dependent Claim Present and Fee Not Previously Paid | | | | | $130 | | $260 |
| | | | | TOTAL | $ $612.00 | | $______ |

☐ Applicant(s) hereby petition for a month extension of time to respond to the outstanding Office Action.

☐ Applicant(s) believe that no (additional) Extension of Time is required; however, if it is determined that such an extension is required, Applicant(s) hereby petition that such an extension be granted and authorize the Commissioner to charge the required fees for an Extension of Time under 37 CFR 1.136 to Deposit Account No. 50-1351.

☒ Enclosed is our Check No. 192 in the amount of $612.00 to cover the additional claim fee and/or extension of time fees.

☒ If the required fees are missing or any additional fees are required to facilitate filing the enclosed response, please charge such fees or credit any overpayment to Deposit Account No. 50-1351 (Order No. SRI1P037A).

Respectfully submitted,

Kevin J. Zilka
Registration No. 41,429

P.O. Box 721030
San Jose, CA 95172
Telephone: (408) 505-5100

(Revised 1/96)

Attorney Docket No.: SRI1P037A (US4116-3)

#8
LSS
12-10-00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION SERIAL NO.: 09/524,095
INVENTOR: Christine Halverson
ASSIGNEE: SRI International
TITLE: Navigating Network-Based Electronic Information Using Spoken Natural Language Input With Multimodal Error Feedback
FILING DATE: March 13, 2000

---

REVOCATION AND POWER OF ATTORNEY

Assistant Commissioner for Patents
Washington, DC 20231

The undersigned assignee of the above-referenced patent application hereby revokes all prior powers of attorney and appoints as his attorney, with full powers of substitution and revocation, to transact all business in the Patent and Trademark Office connected with this application and any patent resulting therefrom, the following:

L. Keith Stephens, Reg. No. 32,632
C. Douglas McDonald, Reg. No. 26,659
John C. Clark, Reg. No. 43,552

Please direct all future communications and telephone calls to:

L. Keith Stephens
CARLTON, FIELDS, WARD, EMMANUEL, SMITH & CUTLER, P.A.
P.O. Box 3239
Tampa, FL 33601-3239
(813) 223-7000

SRI INTERNATIONAL

Date: 11/20/00

By: ______________________
Edward E. Davis, Assistant Secretary

2155

CARLTON FIELDS

ATTORNEYS AT LAW

ONE HARBOUR PLACE
777 S. HARBOUR ISLAND BOULEVARD
TAMPA, FLORIDA 33602-5799

MAILING ADDRESS:
P.O. BOX 3239, TAMPA, FL 33601-3239
TEL (813) 223-7000 FAX (813) 229-4133
Writer's Direct Dial: (813) 229-4209

O I P E
DEC 04 2000
JC17
PATENT & TRADEMARK OFFICE

RECEIVED
DEC 08 2000
Technology Center 2100

November 27, 2000

Assistant Commissioner for Patents
Washington, DC 20231

Re: Patent Application Serial No.: 09/524,095
Inventor: Douglas E. Appelt, et al.
Title: Navigating Network-Based Electronic Information Using Spoken Natural Language Input with Multimodal Error Feedback
Filed: March 13, 2000
Our File No.: 44454/02742

Dear Sir:

Please enter the enclosed Revocation and Power of Attorney into the file of the referenced application.

Very truly yours,

L. Keith Stephens, Reg. No. 32,632

CDM/cm
Enclosure
cc: Edward E. Davis, Asst. Secretary (w/o encl.)

CERTIFICATE OF MAILING

I do hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail, postage prepaid, in an envelope addressed to Assistant Commissioner for Patents, Washington, DC 20231, on the date set forth below.

Cynthia Mejias

11/27/00
Date