BT Technology Journal

Vol.14 No.1 JANUARY 1996

THEME — Speech technology for telecommunications

BT Laboratories
Martlesham Heath Ipswich Suffolk England IP5 7RE

Distributed by Chapman and Hall

The BT Technology Journal is a quarterly periodical of technical papers published by British Telecommunications plc to promote an awareness, among workers in similar fields world-wide, of the Research and Development undertaken by BT in telecommunications and related sciences.

Enquiries to the Editor: (01473) 623232, facsimile (01473) 620915, e-mail: bttj@ipswich.sac.co.uk
Internet access to the BT Laboratories information pages is available on: http://www.labs.bt.com

The BT Technology Journal is distributed by Chapman and Hall, 2-6 Boundary Row, London SE1 8HN, UK.

The Journal is published four times per year in January, April, July and October. Subscription prices for 1996 are: print + Internet access: $214 (USA/Canada) £126 (EU) £140 (all other countries); print only: $185 (USA/Canada) £108 (EU) £122 (all other countries). Subscription prices for individuals are (print only): $95 (USA/Canada) £54 (EU) £54 (all other countries). Individual subscriptions must be paid for by personal cheque or credit card.

Any payment in US$ should be made to Routledge, Chapman and Hall Dollar Account: 051-70700-4, Barclays Bank New York Ltd., 300 Park Avenue, New York, NY 10022, USA.

Subscription rates to USA include airfreight to New York and second class postage thereafter. All other territories outside UK and Europe will be served by accelerated surface post.

Second class postage paid at Rahway, NJ. Postmaster: send address corrections to The BT Technology Journal, c/o Mercury Airfreight International Ltd Inc., 2323 Randolph Avenue, Avenel, NJ 07001, USA (US mailing agent).

All subscription enquiries should be made to Chapman and Hall Subscriptions Department.

All enquiries concerning editorial matters should be made to the Editor, SAC Technographic Ltd, 38 Anson Road, Martlesham Heath, Ipswich, Suffolk IP5 7RG (for voice, fax, e-mail details, see above).

BT Laboratories

Martlesham Heath Ipswich Suffolk England

BT Technology Journal

Vol.14 No.1 JANUARY 1996

THEME Speech technology for telecommunications

Foreword by C Wheddon

Editorial by F A Westall, R D Johnston and A V Lewis

Spoken language systems — beyond prompt and response

P J Wyard, A D Simons, S Appleby, E Kaneen, S H Williams and K R Preston

Spoken language systems allow users to interact with computers by speaking to them. This paper focuses on the most advanced systems, which seek to allow as natural a style of interaction as possible. Specifically this means the use of continuous speech recognition — natural language understanding to interpret the utterance, and an intelligent dialogue manager which allows a flexible style of 'conversation' between computer and user. This paper discusses the architecture of spoken language systems and the components of which they are made, and describes both a variety of possible approaches and the particular design decisions made in some systems developed at BT Laboratories. Three spoken language systems in the course of development are described — a multimodal interface to the BT Business Catalogue, an e-mail secretary which can be consulted over the telephone network, and a multimodal system to allow selection of films in the interactive TV environment.

1. Introduction

No science fiction image of the future is complete without the ever-present personable computer which can understand every word said to them. In spite of these popular media images, the goal of completely natural interaction between humans and machines is still some way off.

Interactive voice response (IVR) systems, which provide services over the telephone network, have been available since the mid-1980s. Initially they were restricted to interactive TouchTone® input with voice providing the response to the user. The use of such services was therefore limited to the population with TouchTone keypads. More recently applications using automatic speech recognition (ASR) have been developed. These often simply allow the option of spoken digit recognition as an alternative to keypad entry, thus allowing the service to be launched even in areas where TouchTone penetration is poor. Moving on from such systems the words which are spoken can be matched to the service. This allows these ASR-based services to be more user-friendly than their TouchTone counterparts because the user can directly answer the question: 'Which service do you require?' with 'weather' or 'sport' rather than 'for weather press 1 for sport press 2', etc. However, they still rely on selection from a predetermined menu of items at any point in the dialogue.

More sophisticated services are now becoming possible using emerging larger vocabulary speech recognition technology. However, it is not sensible to simply extend the menu-based approach to accommodate larger vocabularies. Although well-engineered simple applications may be easy to use, more advanced services are likely to have complicated menu structures. If information can only be provided one item at a time, using a 'prompt and response' dialogue, rigid interaction styles may steer the user through a complex dialogue. This can result in the user becoming lost, or ending up with the wrong information. These problems are particularly significant for inexperienced users. On the other hand, experienced users may become bored by the large number of responses needed when they know exactly what they want. The menu-based structure required by systems which rely on isolated word input is often the limiting factor for new services. This limitation of the user interface is one of the greatest barriers to the usability of many IVR services.

Moving beyond the menu-style interaction towards conversational spoken language will allow users to express their requirements more directly and avoid tedious navigation through menus. This approach will also allow the user to take control of the interaction rather than using the more common 'prompt and response' dialogue.

BT is interested in the development of spoken language systems (SLS) to provide a key competitive advantage. SLSs allow users to interact with computers using conversational language rather than simply responding to system prompts with short or one word utterances. With the rapid increase in competition, service differentiation becomes a key factor in gaining market share. Systems

provide a very useful service for people who are in different time zones, or away from their office, or who need information immediately during unsocial hours. SLSs can be used to automate such services and also those which currently require human operators, thus freeing their time to deal with difficult situations where more complex, or more personalised advice is needed.

Current trends in information networking and the phenomenal growth of the Internet bring their attendant problems for our customers in keeping up with technology, finding what they need, and using information to their best advantage. Spoken language system technology can greatly enhance our customers' ease of access to information, thus increasing network revenue through new and increased usage. Systems which combine several modes of input and output, such as speech, graphics, text, video, mouse-control, touch and virtual reality, are known as multimodal spoken language systems. These allow far greater freedom of expression for users who, as a result, should feel more comfortable and less as though they are 'talking to a computer'. They are able to point, use gestures, speak, type; whatever comes most naturally to them. Spoken language systems will become increasingly important in the near future as progress in technology becomes more widely available.

The goal is to be able to build systems which are not restricted only to those motivated users who are prepared to spend time learning the language the machine understands. These new systems can be used by anyone who wants occasional access to a particular service. They will also help the user successfully gain the information or service they require by simply calling a number and asking for what they want. In fact, the aim is to put back some of the intelligence which existed in the network 50 years ago when a user simply lifted the handset and asked to be connected to the service or number required.

This paper discusses the design and implementation of spoken language systems and is organised as follows. Section 2 gives an outline of the architecture of an SLS. Section 4 describes the components of an SLS in some detail, giving concrete examples from current systems. Section 3 discusses some of the systems currently under development at BTL. These include a multi-modal system for access to the BT Business Catalogue, a speech-in/speech-out system for remote e-mail access and a system for accessing information about films. Section 5 discusses future work which needs to be carried out to improve the quality and usability of SLSs, and section 6 draws some conclusions.

This section outlines a typical spoken language system architecture, from the information processing point of view (platform and inter-process communication issues are not dealt with to any great extent in this paper). The architecture and the key processing components are outlined.

The most basic form of SLS, a speech-in/speech-out (rather than multimodal) system, requires at least the following major components (described briefly below and in more detail in section 4).

- Speech recognition — to convert an input speech utterance to a string of words.
- Meaning extraction — 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.
- Database query — to retrieve the information specified by the output of the meaning extraction component. Some applications (e.g. home banking) may require a specific transaction to occur. Many applications may be a mixture of database query and transaction processing.
- Dialogue manager — this controls the interaction or 'dialogue' between the system and the user, and co-ordinates the operation of all the other system components. It uses a dialogue model (generic information about how conversations progress) to aid the final interpretation of an utterance. This may not have been achieved by the 'meaning extraction' component, because the interpretation relies on an understanding of the conversation as a whole.
- Response generation — to generate the text to be output in spoken form. Information retrieved by the database query component will be passed to the response generation component, together with instructions from the dialogue manager about how to generate the text (e.g. terse/verbose, polite/curt, etc).
- Speech output module (text-to-speech synthesis or recorded speech).

At its simplest, processing consists of a linear sequence of calls to each component, as shown in Fig 1. A typical output of each stage from an application which accesses the BT Business Catalogue is shown. It is not necessary to understand the output of the 'meaning extraction' component in detail to realise that meaning extraction can be a non-trivial exercise. The simple linear sequence shown in Fig 1 is, in general, too inflexible. It is better if the dialogue manager is given greater control, to call the other components in a flexible order, according to the results at

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.