



РЕПУБЛИКА СРБИЈА
ЗАВОД ЗА ИНТЕЛЕКТУАЛНУ СВОЈИНУ

Број: 2380/07 П-551/06
Београд, 21.03.07

О Б А В Е Ш Т Е Њ Е

У спроведеном поступку формалног испитивања, утврђено је да пријава, сходно одредби члана 40. став 1. Закона о патентима ("Службени лист СЦГ", бр. 32/04), испуњава услове за објаву.

Објавити у "Гласнику интелектуалне својине" број
следеће податке о пријави патента:

2 - 2007

04 JUN 2007

(51) МКП⁽⁷⁾: Н 04 В 3/20

(11) Број документа: П- 551/06

(13) А

(21) Број пријаве: П- 551/06

(22) Датум подношења: 04.10.06

(61) Број основне пријаве: П-
или патента:

(62) Број првобитне пријаве: П –

(30) Подаци о праву првенства:

Земља: РС

Датум: 04.10.2006

Број: П-551/06

(23) Датум излагања на међународној изложби:

(54) Назив проналаска: Систем и поступак са слободну говорну комуникацију помоћу
микрофонског низа

(YU)

(EN) System and technique for hands-free voice communication using microphone array

(71) Подносилац пријаве: MicronasNIT, Фрушкогорска 11а Нови Сад

(72) Проналазач-и ДР Зоран Шарић, Вукасовићева 65/7 Нови Београд; Др Слободан Јовичић
Вишњички венац 67, Београд; Др Владимир Ковачевић Радничка 35 А; Др Никола Теслић Бул. Цара
Лазара 29; Др Драган Кукољ, Народног фронта 31; Нови Сад

(74) Пуномоћник:

Позива се подносилац пријаве да, уколико је у могућности, Заводу достави текст патентне пријаве, превод на енглески језик назива проналаска и апстракта у електронској форми, чиме ће се знатно убрзати техничке припреме за објаву предметне пријаве патента.

На захтев подносиоца пријаве, Завод ће извршити суштинско испитивање услова патентбилности. Наведени захтев подноси се након објаве пријаве патента у "Гласнику интелектуалне својине", а *најкасније у року од 6 месеци од дана објаве*. У случају пропуштања напред наведеног рока, подносилац пријаве може поднети захтев за суштинско испитивање услова патентбилности у накнадном року од 30 дана од дана пријема обавештења о протеклу рока. О дану објаве пријаве патента, подносилац пријаве ће бити накнадно обавештен, посебним дописом Завода.

Уз захтев, подносилац је дужан доставити доказ о уплаћеној такси у износу од **3600,00 динара**. Такса се уплаћује на жиро рачун број 840-742221843-57, са позивом на број 97, шифра са контролним бројем општине седишта уплатиоца (видети Прилог 3 Правилника о условима и начину вођења рачуна за уплату јавних прихода и распоред средстава са тих рачуна, "Службени гласник РС", бр. 20/2007 од 20.02.2007 године), са назнаком "Такса за суштинско испитивање" и уз навођење броја пријаве патента (прималац: Републичке административне таксе). Поднети захтев се не може повући.

Подносилац пријаве који је истовремено и домаћи проналазач, наведени таксу плаћа у износу од 10%.

Ако подносилац пријаве у наведеном року не поднесе захтев за суштинско испитивање услова патентбилности пријаве за признање патента, пријава за признање патента ће се сматрати повученом, што ће се утврдити закључком.

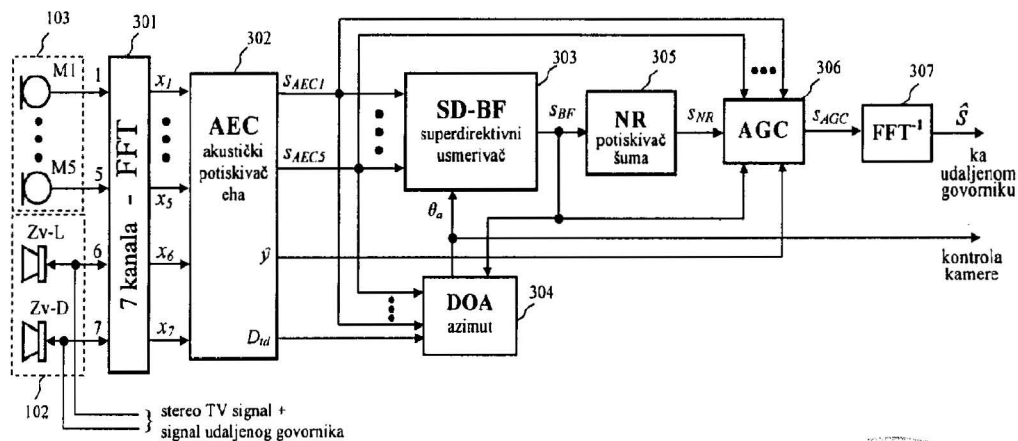
Обавештење доставити:

- подносиоцу пријаве, путем заступника
MicronasNIT Фрушкогорска 11А
Нови Сад
- Регистру, АОП
- У спис



APSTRAKT

Pronalazak se odnosi na sistem i postupak za slobodnu govornu komunikaciju u video-telefonskim ili telekonferencijskim primenama zasnovan na mikrofonskom nizu čiji je cilj kvalitetno snimanje govornika u prostoriji u uslovima većeg rastojanja, prisutnih smetnji, akustičkog eha od udaljenog govornika i TV programa, reverberacije prostorije i kretanja govornika u prostoriji. Sistem se sastoji od: digitalnog TV prijemnika i digitalne kamere za reprodukciju i snimanje slike, respektivno, stereo zvučnika i mikrofonskog niza za reprodukciju i snimanje zvuka, respektivno, pojačavačkog i akvizicionog modula za audio signale i DSP za obradu akustičkih signala. Postupak za obradu mikrofonskih signala se izvršava u frekvencijskom domenu i sastoji se od: potiskivanja akustičkog eha nastalog od signala udaljenog govornika i stereo signala TV programa, prostornog filtriranja bliskog govornika u odnosu na izvore smetnji i reverberaciju prostorije na bazi adaptivne karakteristike usmerenosti mikrofonskog niza, lociranja govornika u horizontalnoj ravni, potiskivanja svih rezidualnih smetnji i adaptivne kontrole pojačanja predajnog signala.



SISTEM I POSTUPAK ZA SLOBODNU GOVORNU KOMUNIKACIJU POMOĆU MIKROFONSKOG NIZA

OBLAST TEHNIKE NA KOJU SE PRONALAZAK ODNOSI

Pronalazak pripada oblasti obrade akustičkog signala, ili konkretnije, metodama poništavanja akustičkog eha, prostornog selektovanja i lociranja govornika u reverberantnom akustičkom ambijentu i potiskivanja šuma primenom mikrofonskog niza.

TEHNIČKI PROBLEM

Slobodni, "hands-free" (engl.), komunikacioni sistemi za prenos govornog signala u punom dupleksu koriste se u mnogim aplikacijama kao što su: video-telefonski sistemi, telekonferencijski sistemi, spikerfoni u prostoriji ili kolima, komunikacija čovek-računar putem glasa, itd. "Hands-free" govorna komunikacija podrazumeva da se govornik nalazi u akustičkom ambijentu na određenoj distanci od interfejsnih elemenata sistema - mikrofona i zvučnika. Ovakvi uslovi odvijanja govorne komunikacije generišu više tehničkih problema koje je potrebno rešiti da bi se održao kvalitet komunikacije na prihvatljivom nivou.

Osnovni problem jeste akustički eho koji nastaje prenosom dela akustičke energije iz zvučnika u mikrofona tako da sagovornik na udaljenom kraju čuje sopstveni glas kao smetnju. Konvencionalno, poništavanje signala eha obavlja adaptivni filter estimiranjem prenosne funkcije akustičkog puta između zvučnika i mikrofona, tako da se na njegovom izlazu dobija približno isti signal kao što je signal akustičkog eha. Oduzimanja ova dva signala se poništava akustički eho. Međutim, poništavanje eha ne može biti idealno zbog nelinearnosti sistema i nestacionarnosti akustičkog ambijenta. Kao rezultat pojavljuje se rezidualni signal eha. Pri tome ostaje osnovni zahtev da snimljeni govorni signal na bližem kraju ne sme biti izobličen primenom postupka potiskivanja eha.

U akustičkom ambijentu akustičke smetnje mogu biti različite prirode i uzroka. One mogu biti stacionarne i nestacionarne (na primer kao što su šum računara ili buka u

automobilu) i poticati od više izvora lociranih na različitim pozicijama u prostoru gde se nalazi govornik. Pored toga u zatvorenim prostorima (radne sobe, sale, automobilska kabina) pojavljuje se efekat reverberacije, koji se manifestuje kao difuzna smetnja. Pošto se govornik najčešće nalazi u ovakvom ambijentu onda se mora izvršiti njegova separacija od ostalih izvora smetnji kako bi se omogućilo samo njegovo snimanje. Konvencionalno, ovaj problem se rešava primenom mikrofonskog niza koji se sastoji od više mikrofona poređanih na minimalnoj međusobnoj distanci. Određena konfiguracija mikrofona omogućava dobijanje sistema sa usmerenom karakteristikom osetljivosti. Ovakav mikrofonski sistem ima dovoljno uzanu karakteristiku usmerenosti da u prostoru ambijenta može snimiti samo odabranog govornika dok ostale izvore smetnji koji se nalaze na drugim pozicijama (lokacijama) može potisnuti i time ostvariti dobitak u odnosu izabrani govornik – ostale smetnje. Veličina ovog dobitka zavisi od: karakteristike usmerenosti mikrofonskog niza (širine osnovne petlje), veličine bočnih petlji, separabilnosti govornika i izvora smetnji (da nisu suviše blizu), veličine reverberacije, nestacionarnosti svih izvora signala, itd.

Određivanje pravca u prostoru na kome se nalazi izabrani govornik i usmeravanje karakteristike usmerenosti mikrofonskog niza ka njemu jeste važan problem u “hands-free” komunikacionim sistemima. Postupci određivanja pravca su veoma osetljivi na sve smetnje prisutne u ambijentu i posebno: na nestacionarnost izabranog govornika (kada se on kreće u ambijentu) i kada se u datom ambijentu nalazi više govornika koji istovremeno govore (*cocktail-party* efekat). Određivanje pravca aktuelnog govornika u odnosu na mikrofonski niz u horizontalnoj ravni je veoma važno u video-telefonskim i telekonferencijskim sistemima, jer je neophodno odrediti koordinate za kontrolu video kamere.

Kod snimanja govora u akustičkom ambijentu uvek se pojavljuje problem aditivnog stacionarnog i/ili nestacionarnog šuma kao i rezidualnog šuma u obradi akustičkog signala. Ovi šumovi degradiraju kvalitet snimljenog govornog signala a ukoliko su dovoljno intenzivni mogu izazvati i narušavanje njegove razumljivosti. Postoji mnogo algoritama za potiskivanje šuma, optimiziranih za pojedine vrste šumova, ali se uvek postavlja zahtev da se ostvari određen dobitak u poboljšanju odnosa signal/šum pod uslovom da se ne unesu izobličenja u govorni signal i time dodatno ne naruši njegova razumljivost.

Promenljivi ambijentalni uslovi i posebno, promenljivo rastojanje govornik-mikrofonski niz, zahtevaju automatsku kontrolu pojačanja sistema kako bi nivo glasa

govornika bio što stabilniji i prijatniji za slušaoca na udaljenom kraju telekomunikacionog kanala. Automatska kontrola pojačanja u sistemima koji rade u punom dupleksu zahteva dodatne informacije od detektora aktivnosti govora na bližem kraju, detektora aktivnosti govora na daljem kraju kao i potiskivača akustičkog eha.

Iz izloženog se vidi da su tehnički problemi u rešenju slobodnog, "hands-free", komunikacionog sistema za prenos govornog signala u punom dupleksu i njegovu primenu u video-telefonskim i/ili telekonferencijskim sistemima veoma složeni i da zahtevaju integralni pristup u optimizaciji rešenja, posebno kada se ima u vidu rad sistema u realnom vremenu na bazi komercijalne platforme digitalnog procesora signala (DSP).

STANJE TEHNIKE

Kvalitetno snimanje govora u uslovima prisustva akustičkih smetnji i reverberacije prostorije predstavlja složen problem. U uslovima kada se spektri korisnog govornog signala preklapaju sa spektrima prisutnih smetnji, jednokanalnim postupcima obrade nije moguće ostvariti značajnije poboljšanje kvaliteta govornog signala. Sa razvojem digitalne obrade signala i postizanjem dovoljno velike računarske snage DSP-a otvoren je put za primenu višemikrofonskih postupaka obrade akustičkih signala. Prednost mikrofonskih nizova u odnosu na jednokanalne postupke obrade je njihova sposobnost da prilagode svoju prostornu karakteristiku prijema (karakteristiku usmerenosti) trenutnom prostornom rasporedu odabranog govornika i smetnji. Pri tome ostvaruju maksimalno potiskivanje prisutnih smetnji uz istovremeno isticanje odabranog govornika. Osnovni problemi koji se u primeni mikrofonskih nizova sreću su sledeći (M.S. Brandstein, D.B. Ward (Eds.), *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin 2001; Y. Huang, J. Benesty, *Audio signal processing for next generation multimedia communication systems*, Kluwer Academic Publishers Publ., 2004.): nepoznavanje tačne lokacije odabranog govornika, nepoznavanje broja i prostornog rasporeda prisutnih smetnji, višestruke refleksije korisnog izvora i smetnji o zidove prostorije i nestacionarnost izvora akustičkih smetnji i odabranog govornika.

Kada se mikrofonski niz upotrebi u video-telefonskim ili telekonferencijskim sistemima koji funkcionišu u punom dupleksu, onda se broj problema uvećava. Najveći problem je pojava akustičkog eha, zatim potreba za automatskom regulacijom pojačanja

(AGC) predajnog dela sistema, kao i moguća pojava nestabilnosti sistema, tzv. mikrofonija. Dodatni problem koji ovaj patent razmatra je postojanje signala TV programa koji se kao aditivni akustički eho pojavljuje na ulazu mikrofonskog niza.

Veliki broj navedenih problema generisao je veoma različita rešenja koja su patentirana i koja rešavaju ili pojedinačne probleme ili integralno nekoliko problema. Na primer: U.S. objavljena patentna prijava 2006/0153360 A1, prijavljen 2. septembra 2005., sa naslovom „Speech signal processing with combined noise reduction and echo compensation“, daje integralno rešenje potiskivača eha i potiskivača šuma, zatim U.S. patent 7,035,415 B2, prijavljen 15. maja 2001, sa naslovom „Method and device for acoustic echo cancellation combined with adaptive beamforming“, koji daje integralno rešenje potiskivača eha i rešenje za formiranje usmerene karakteristike mikrofonskog niza, zatim EP objavljena patentna prijava 1 633 121 A1, prijavljen 3. septembra 2004., sa naslovom „Speech signal processing with combined adaptive noise reduction and adaptive echo compensation“, daje integralno rešenje potiskivača rezidualnog eha i potiskivača šuma, zatim EP objavljena patentna prijava 1 571 875 A2, prijavljen 23. februara 2005., sa naslovom „A system and method for beamforming using a microphone array“, koji daje rešenje samo za formiranje usmerene karakteristike mikrofonskog niza, zatim EP objavljena patentna prijava 1 581 026 A1, prijavljen 17. marta 2004., sa naslovom „Method for detecting and reducing noise from a microphone array“, daje rešenje samo za potiskivanje šuma u mikrofonskom nizu, kao i EP objavljena patentna prijava 1 286 175 A2, prijavljen 1. avgusta 2002., sa naslovom „Robust talker localization in reverberant environment“, daje rešenje samo za lokalizaciju govornika u reverberantnoj sobi.

Integralno rešenje svih naznačenih problema, izloženo u ovom patentu, objedinjuje pozitivne osobine pojedinih postupaka obrade signala u rešenju svakog od naznačenih problema, integralno ih rešava u frekvencijskom domenu optimizirajući računarske resurse i daje rešenje koje u realnom vremenu obezbeđuje kvalitetnu slobodnu govornu komunikaciju u video-telefonskim i/ili telekonferencijskim sistemima.

IZLAGANJE SUŠTINE PRONALASKA

Predmet ovog pronalaska je sistem za slobodnu govornu komunikaciju u video-telefonskim ili telekonferencijskim primenama koji koristi mikrofonski niz i složenu

obradu akustičkog signala u cilju obezbeđenja kvaliteta i razumljivosti govornog signala u složenom akustičkom ambijentu i u kome su mnogi prethodno nabrojani nedostaci pojedinačno ili integralno eliminisani.

Sistemom, koji je predmet pronalaska, prenosi se govor a kao prenosni medijum se koristi digitalna televizija. Za snimanje i reprodukciju govornog signala koristi se mikrofonski niz i zvučnici, respektivno, koji su sastavni elementi TV prijemnika. Pošto je reč o video-telefonskim ili telekonferencijskim primenama, za snimanje i reprodukciju slike koristi se digitalna kamera i digitalni TV prijemnik, respektivno.

Sušтина pronalaska jeste u specifičnoj obradi govornog signala koji se snima u akustičkom ambijentu prostorije u kojoj se nalazi sistem i govornik. Za snimanje govornika u prostoriji, koji se nalazi na određenom rastojanju (do nekoliko metara) od TV prijemnika, sistem koristi mikrofonski niz od N mikrofona. Mikrofonski niz snima sve signale u prostoriji: koristan signal kao direktan talas koji stiže od govornika do mikrofona i signale smetnji koji mogu biti raznovrsni. Kao signali smetnje pojavljuju se: akustički eho kao direktan zvučni talas iz zvučnika preko kojih se emituje glas sagovornika sa udaljenog kraja komunikacionog kanala, akustički eho kao direktan zvučni talas iz zvučnika preko kojih se emituje stereo TV program, direktni talasi od jednog ili više izvora šumova ili izvora drugih smetnji koji se mogu naći u prostoriji i svi reflektovani talasi (eho prostorije) koji potiču od svih izvora zvukova, uključujući i govornika, a koji nastaju usled reverberacije prostorije. Treba naglasiti da izvori zvukova u prostoriji mogu biti stacionarni ili nestacionarni, što je najčešći slučaj, kako po svojim karakteristikama tako i po lokaciji u prostoriji (pokretni izvori zvukova).

Različite smetnje zahtevaju različite tehnike za njihovo eliminisanje i suština pronalaska jeste u optimalnom projektovanju algoritama koji treba da maksimalno eliminišu smetnje i da obezbede najbolji kvalitet govornog signala koji se prenosi do sagovornika na udaljenom kraju komunikacionog kanala.

Mikrofonski signali iz mikrofonskog niza se obrađuju u digitalnoj formi u DSP, kompletno u frekvencijskom domenu. Ovaj domen omogućava određene prednosti u pogledu brzine obrade i broja računskih operacija, što je veoma važno za DSP i rad u realnom vremenu. Za potiskivanje akustičkog eha neophodno je da se u DSP uvedu i signali iz zvučnika.

U DSP-u se izvršava više složenih algoritama: algoritam za potiskivanje signala akustičkog eha (AEC – *Acoustic Echo Cancelling*), algoritam za obradu mikrofonskih signala u cilju formiranja adaptivne karakteristike usmerenosti mikrofonskog niza

(ABF – *Adaptive Beam Forming*), algoritam za ocenu pravca dolaska korisnog signala (DOA – *Direction of Arrival*) odnosno lociranje govornika u prostoriji, algoritam za potiskivanje stacionarnog i nestacionarnog šuma i rezidualnog eha (NR – *Noise Reduction*) i algoritam za automatsku kontrolu pojačanja sistema (AGC – *Automatic Gain Control*) radi kompenzacije različite udaljenosti govornika od mikrofonskog niza. Pored ovih osnovnih algoritama u DSP-u se izvršava i više drugih algoritama kao što su: detektor aktivnosti govora (VAD – *Voice Activated Detector*) na bližem kraju, VAD na daljem kraju, detektor istovremene aktivnosti govora na oba kraja (DTD – *Double Talk Detector*), dodatno filtriranje radi redukcije šuma (PF – *Post Filtering*), itd. Cilj svih navedenih algoritama je maksimalna redukcija svih smetnji uz minimalnu degradaciju govornog signala i time obezbeđivanja maksimalnog kvaliteta predajnog govornog signala.

Specifičan aspekt pronalaska se nalazi u adaptivnom potiskivanju akustičkog eha pomoću adaptivnih filtara koji modeliraju prenosnu karakteristiku akustičkog puta od zvučnika do mikrofona. Prenosna karakteristika je složena jer se radi o prenosnom putu od 2 (stereo) zvučnika do N mikrofona u mikrofonskom nizu, zbog čega se svaki mikrofonski signal filtrira sopstvenim adaptivnim filtrom. Kontrolu rada adaptivnih filtara vrši detektor aktivnosti govora na oba kraja.

Sledeću specifičnost pronalaska čini adaptivna karakteristika usmerenosti mikrofonskog niza koja omogućava prostorno filtriranje, odnosno izdvajanje pravca u prostoru na kome se nalazi govornik i gde se koristan signal maksimalno pojačava u odnosu na signale iz ostalih pravaca koji se slabe. Usmerena karakteristika mikrofonskog niza se ostvaruje adaptivnim ponderisanjem i sumiranjem mikrofonskih signala, što obezbeđuje stabilan indeks usmerenosti u frekvencijskom domenu i veću robusnost sistema za slobodnu govornu komunikaciju u reverberantnom akustičkom ambijentu.

Određivanje dolaznog pravca direktnog akustičkog talasa od govornika je naredna specifičnost pronalaska. Ova funkcija u sistemu slobodne govorne komunikacije je neophodna za kontrolu i upravljanje usmerenom karakteristikom mikrofonskog niza po azimutu, a može se koristiti i za kontrolu i upravljanje video kamere. Ona koristi mikrofonske signale posle potiskivanja akustičkog eha. Nakon određivanja generalizovane kroskorelacije mikrofonskih signala i njihovih faznih transformacija, estimira se dolazni pravac direktnog akustičkog talasa govornika. Ova funkcija je pod direktnom kontrolom detektora aktivnosti govora.

Sljedeću specifičnost pronalaska čini postupak adaptivnog potiskivanja stacionarnog i nestacionarnog šuma. Postupak je realizovan na bazi nelinearnog kompresora estimiranog šuma koji se određuje u nekoliko podopsega. Koriste se dve estimacije šuma koje obezbeđuju rezultat potiskivanja optimiziran prema karakteristikama govornog signala. To je učinjeno iz razloga potrebe da proces adaptivnog potiskivanja šuma ne sme degradirati govorni signal. Proces filtriranja se završava adaptivnim Wiener-ovim post-filtrom.

Specifičan aspekt pronalaska jeste i automatska kontrola pojačanja govornog signala pre predaje ka udaljenom sagovorniku. Ova specifičnost je važan sastavni elemenat sistema za slobodnu govornu komunikaciju. Sistem obezbeđuje kompenzaciju različitih intenziteta govornog signala, kao individualnih karakteristika govornika, ali i različite intenzitete govora u zavisnosti da li se govornik nalazi bliže ili dalje u odnosu na mikrofonski niz. Rešenje pravi razliku da li je govornik aktivan ili se u korisnom signalu pojavljuje: pauza, rezidualni cho, akustička smetnja ili signal govora sa udaljenog kraja; zbog toga rešenje koristi više informacija prethodno detektovanih u sistemu. Analiza mogućeg scenarija mora biti pouzdana, u protivnom može doći do negativnog efekta slabljenja korisnog govornog signala.

Inventivnost u ovom pronalasku se nalazi u poboljšanju svake od navedenih specifičnosti, ali i u postupku integrisanja svih algoritama u jedinstvenu celinu koja funkcioniše stabilno i kvalitetno. Algoritamske procedure su optimizirane korišćenjem zajedničkih resursa.

Ovi i drugi aspekti, specifičnosti i benefiti ovog pronalaska biće očigledniji nakon uvida u detaljan opis pronalaska, patentne zahteve i pripadajuće crteže.

KRATAK OPIS SLIKA I NACRTA

Slika 1 – prikazuje elemente sistema za slobodnu video-telefonsku komunikaciju pomoću mikrofonskog niza i digitalne televizije.

Slika 2 – prikazuje ambijentalne uslove primene sistema za slobodnu video-telefonsku komunikaciju pomoću mikrofonskog niza.

Slika 3 – prikazuje blok dijagram podsistema za obradu audio signala u okviru sistema za slobodnu video-telefonsku komunikaciju; on sadrži mikrofonski niz sa adaptivnom karakteristikom usmerenosti (SD-BF), blok za lociranje govornika u

prostoru (DOA), blok za potiskivanje eha (AEC), blok za potiskivanje šuma (NR) i blok za automatsku kontrolu pojačanja (AGC).

Slika 4 – prikazuje blok dijagram za potiskivanje akustičkog eha (AEC).

Slika 5 – prikazuje blok dijagram za adaptivno određivanje pravca bliskog govornika po horizontali (DOA-azimut).

Slika 6 – prikazuje blok dijagram za prostorno filtriranje (SD-BF).

Slika 7 – prikazuje blok dijagram za potiskivanje šuma (NR).

Slika 8 – prikazuje blok dijagram za automatsku regulaciju pojačanja (AGC).

DETALJAN OPIS PRONALASKA

Ovaj pronalazak opisuje sistem i postupak obrade akustičkog signala za slobodnu govornu komunikaciju pomoću mikrofonskog niza.

Slika 1 prikazuje elemente sistema za slobodnu video-telefonsku komunikaciju pomoću mikrofonskog niza i digitalne televizije. Digitalni televizor **100**, koji korisniku normalno služi za praćenje TV programa, u sistemu za slobodnu video-telefonsku komunikaciju koristi se kao video monitor za video komunikaciju sa sagovornikom i kao audio terminal za audio komunikaciju. Naime, kada se putem komunikacionog kanala **101** dobije poziv i uspostavi veza sa sagovornikom tada se televizor **100** koristi kao multimedijalni interfejs gde se preko zvučnika **102** sluša sagovornik a na delu ekrana **105** televizora **100** prati se slika sagovornika. Istovremeno, na udaljenom kraju komunikacionog kanala, sagovornik na sličnom TV prijemniku vidi sagovornika sa bližeg kraja, koga snima kamera **104** i mikrofonski niz **103**. Kamera **104** je pokretna i njom se upravlja na bazi koordinata koje se dobijaju obradom mikrofonskih signala iz mikrofonskog niza **103**.

Analogni signali iz mikrofona u mikrofonskom nizu **103** se pojačavaju pomoću pojačavača **106** i zajedno sa stereo signalima iz zvučnika **102** se uvode u akvizicioni modul **107**, gde se digitalizuju i tako digitalizovani predaju DSP-u **108** na dalju obradu. Obradeni govorni signal govornika na bližem kraju pomoću DSP-a **108** prenosi se preko komunikacionog kanala **101** do sagovornika na daljem kraju. Obradom akustičkih signala u DSP-u **108** dobijaju se prostorne koordinate lociranja govornika u prostoriji u kojoj se nalazi sistem za slobodnu komunikaciju, pomoću kojih DSP **108** upravlja sa

pokretnom kamerom **104** usmeravajuću je ka govorniku. Na taj način se ostvaruje potpuno slobodna audio i video komunikacija dva sagovornika preko sistema digitalne televizije.

Slika 2 šematski prikazuje ambijentalne uslove primene sistema za slobodnu video-telefonsku komunikaciju pomoću mikrofonskog niza; prikazan je samo deo sistema koji se odnosi na obradu akustičkog signala. U prostoriji **201** nalaze se sistem za slobodnu video-telefonsku komunikaciju, govornik **202** i izvor šuma **203**, što je uobičajeno za svaki akustički ambijent. Preko zvučnika **102** stereo audio sistema digitalne televizije govornik **202** sluša dolazni govorni signal **204** sagovornika sa udaljenog kraja najčešće kao mono signal. Zvuk u ambijentu prostorije **201** snima mikrofonski niz **103** sastavljen od *N* mikrofona. Nakon kompleksne obrade mikrofonskih signala u bloku **207** govorni signal govornika **202** se preko bloka **208** prenosi ka udaljenom sagovorniku kao mono signal.

Ambijentalni uslovi odvijanja govorne komunikacije u prostoriji **201** su veoma kompleksni. Kod slobodne video-telefonske komunikacije u prostoriji **201** postoji minimum tri izvora zvuka: stereo zvučnici **102** koji emituju govor udaljenog sagovornika i TV program, govornik **202** i bar jedan izvor šuma **203**. U prostoriji može biti i više izvora šumova: šum računara, šum klima sistema, buka sa ulice koja prodire u prostoriju kroz prozore, buka iz susednih prostorija, vibracije zgrade, ili drugi govornik, više govornika, izvor muzike, itd. Dakle, pojavljuje se veoma složena akustička slika u prostoriji. Mikrofonski niz **103** snima, kao senzorski sistem, sve zvuke u prostoriji, snima direktne zvučne talase od svakog izvora ali i sve refleksije od zidova prostorije i drugih predmeta koji se nalaze u njoj. Tako na primer, od zvučnika **102** do mikrofonskog niza **103** stiže direktan talas **209** i mnogi reflektovani talasi od kojih je samo jedan **210** prikazan na slici 2; od govornika **202** stiže direktan talas **211** i pored ostalih i dva reflektovana talasa **212a** i **212b**, od izvora šuma **203** stiže direktan talas **213** i pored ostalih i reflektovani talas **214**.

Od svih zvukova koje mikrofonski niz snima jedino je direktan talas **211** od govornika **202** koristan signal, svi ostali su smetnje. Od svih smetnji najveća je akustički eho **209** koji dolazi iz zvučnika **102**. Sve ostale refleksije zborno čine reverberaciju prostorije. Zadatak bloka za obradu audio signala **207** jeste da potisne signal akustičkog eha, da selektuje koristan signal **211** od svih ostalih smetnji, da potisne signale reverberacije i da potisne direktne signale izvora smetnji, kojih može da bude i više od jednog izvora. Poseban zadatak bloka **211** jeste adaptivno praćenje

nestacionarnosti akustičke scene u prostoriji bilo da se govornik pokreće, ili da se od razgovora do razgovora nalazi na različitim pozicijama u prostoriji, ili da se izvori šumova pokreću, da su nestacionarni ili da menjaju svoje karakteristike. U daljem tekstu biće pojedinačno opisana rešenja koja su u ovom pronalasku primenjena.

Na slici 3 prikazana je blok šema kompletnog postupka obrade audio signala u okviru sistema za slobodnu video-telefonsku komunikaciju pomoću mikrofonskog niza. Svi mikrofonski signali **103**, od M1 do M5, kao i signali stereo zvučnika **102**, Zv-L i Zv-D, se digitalizuju u akvizicionom bloku **107**, slika 1, i konvertuju u frekvencijski domen pomoću brze Fourierove transformacije (FFT) **301** u signale x_1 do x_7 . Treba naglasiti da mikrofonski niz sadrži 5 mikrofona u rešenju ovog patenta, ali se može primeniti veći broj mikrofona ukoliko određena aplikacija to zahteva. U bloku **302** vrši se potiskivanje akustičkog eha u svim signalima x_1 do x_5 , koristeći signale x_6 i x_7 kao referentne. Signali sa potisnutim ehom S_{AEC1} do S_{AEC5} koriste se u bloku **304** za određivanje pravca direktnog zvučnog talasa **DOA** (*Direction Of Arrival*) po horizontali (azimutu θ_a) od aktuelnog govornika i time omogućava njegovo praćenje u prostoriji. Na osnovu ocenjenog ugla θ_a u bloku **303** se optimiziraju težinski koeficijenti signala x_1 do x_5 u cilju formiranja karakteristike horizontalne usmerenosti mikrofonskog niza sa maksimumom prijema na pravcu θ_a . Karakteristika prijema formirana u bloku **303** ima superdirektivno svojstvo što znači da joj je indeks usmerenosti (direktivnosti) prijema veći u odnosu na karakteristiku koja bi se dobila samo kompenzacijom kašnjenja i sumiranjem mikrofonskih signala.

U bloku **303** vrši se vremenska kompenzacija međusobnog kašnjenja akustičkih signala od govornika do mikrofona. Kontrolom ovog kašnjenja signalom DOA (θ_a) iz bloka **304**, omogućava se upravljanje karakteristikom usmerenosti mikrofonskog niza po azimutu. Takođe, u bloku **303** formira se karakteristika usmerenosti mikrofonskog niza, **SD-BF** (*Superdirective Beamformer*). Ova karakteristika ima osnovnu petlju usmerenja dovoljno uzanu i usmerenu u željenom pravcu, dok su bočne petlje znatno manje po intenzitetu. Time se omogućava mikrofonskom nizu prostorno filtriranje, odnosno separaciju izvora zvukova po horizontali. Ovako formirana karakteristika usmerenosti je veoma bitna sa aspekta utišavanja signala bočnih smetnji u odnosu na korisni signal i sa aspekta smanjenja efekta reverberacije prostorije. Karakteristika usmerenosti se formira ponderisanjem mikrofonskih signala i njihovim sumiranjem u jedinstveni izlazni signal.

Signal na izlazu bloka **303** sadrži koristan govorni signal i signal smetnji koji se sastoji od rezidualnog signala nakon potiskivanja akustičkog eha, potisnut šum ambijenta i potisnute signale reverberacije. Ovaj signal ulazi u blok **NR** (*Noise Reduction*) **305** gde se vrši dodatno potiskivanje signala smetnji. Proces potiskivanja je adaptivan obzirom na nestacionarnost signala smetnji. Takođe, važan zahtev u realizaciji **NR** bloka jeste da proces potiskivanja šuma ne sme da utiče na kvalitet govornog signala.

Finalni blok obrade signala u sistemu za slobodnu govornu komunikaciju u video-telefonskim ili telekonferencijskim primenama jeste blok **306** za automatsku kontrolu pojačanja **AGC** (*Automatic Gain Control*) obrađenog govornog signala. U ovom bloku koristi se više informacija iz celokupnog sistema koje su važne za definisanje mogućih uslova u kojima se govorni signal može naći i gde je potrebno na odgovarajući način izvršiti njekovu amplitudsku korekciju. Na taj način se može obezbediti približno isti nivo predajnog govornog signala nezavisno od udaljenosti aktuelni govornik od mikrofonskog niza i obezbediti njegov bolji kvalitet na udaljenom kraju komunikacionog kanala.

Na izlazu sistema rezultat obrade signala se transformiše iz frekvencijskog u vremenski domen pomoću inverzne FFT u bloku **307**. Estimirani govorni signal na bližem kraju ($\hat{s}$) se prenosi kroz kanal ka udaljenom sagovorniku.

Na slici 4 prikazan je blok dijagram potiskivača akustičkog eha (AEC) **302**, koji se sastoji od dva osnovna bloka: blok **401** koji se sastoji od 5 adaptivnih NLMS (*Normalized Least Mean Square*) algoritama i bloka **402** čija je osnovna funkcija detekcija aktivnosti govora bliskog i udaljenog govornika DTD (*Double Talk Detection*).

NLMS algoritmi, NLMS1 do NLMS6, obrađuju signale iz mikrofona x_1 do x_5 i obrađene signale s_{AEC1} do s_{AEC5} prosleđuju dalje ka blokovima **303**, **304** i **306**, slika 3. Funkcija NLMS algoritama je potiskivanje eha u svakom od mikrofonskih signala. Ovu funkciju omogućavaju referentni signali iz zvučnika **102** i kontrolni signali iz DTD detektora **402**. NLMS algoritam modelira prenosnu funkciju akustičkog puta od svakog zvučnika **102** do svakog mikrofona **103**; na primer NLMS1 modelira prenosne funkcije h_{L1} od zvučnika Zv-L do mikrofona M1 i h_{D1} od zvučnika Zv-D do mikrofona M1, itd. Prolaskom signala iz zvučnika kroz NLMS filtre dobija se replika signala na mikrofona koji su došli akustičkim putem i oduzimanjem ova dva signala postiže se

potiskivanje eho signala na izlazu NLMS algoritama. U cilju boljeg potiskivanja eha, kao i u slučaju **RLS1** AEC algoritma (**RLS** – *Recursive Least Squares*) koji se dole opisuje, koriste se DFT koeficijenti iz prethodnih blokova obrade. Kako NLMS algoritam zahteva znatno manje računarskog vremena u odnosu na RLS algoritam, u realizaciji NLMS algoritama se koriste DFT koeficijenti iz prethodna 5 bloka obrade.

Blok **403** sa oznakom **RLS1** AEC je ključni algoritamski deo postupka detekcije dvostruke govorne aktivnosti iz bloka **402**. **RLS1** AEC vrši grubo potiskivanje akustičkih smetnji u signalu iz mikrofona M1 primenom RLS algoritma. RLS algoritam ima brzu konvergenciju što obezbeđuje dobru estimaciju govornog signala kao i estimaciju aditivne komponente eho signala. S obzirom da veličina primenjenog DFT prozora od 1024 nije dovoljno velika da bi se ostvarilo maksimalno potiskivanje eho smetnji u prostoriji sa velikom reverberacijom, regresionom vektoru se pridružuju DFT koeficijenti iz 3 prethodna bloka obrade. Time se ostvaruje dvostruki dobitak: maksimalno potiskivanje eha i kašnjenje signala kroz sistem se ne uvećava jer red DFT ostaje nepromenjen.

Izlaz iz **RLS1** AEC bloka su dva signala e i $\hat{y}$. Prvi signal e je estimacija govora bliskog govornika na mikrofona M1. Drugi signal $\hat{y}$ je estimacija aditivne komponente signala eha u signalu mikrofona M1. Oba ova signala se koriste za detekciju dvostruke govorne aktivnosti koja se realizuje u bloku **402** sa oznakom **DTD**. Signal iz **DTD** detektora kontroliše rad NLMS algoritama u smislu da sprečava adaptaciju algoritama NLMS 1 do NLMS 5 za vreme dvostruke aktivnosti govora, kada dolazi do remećenja rada adaptivnih algoritama. U bloku **405** vrši se usrednjavanje snaga signala na zvučnicima prema relaciji:

$$P_{ref} = \frac{|x_6|^2 + |x_7|^2}{2}. \quad (1)$$

Na oba signala $\hat{y}$ i P_{ref} se primenjuje rekurzivno usrednjavanje, tako da se dobijaju usrednjene snage signala eha u mikrofona M1 **(2)** i signala na zvučnicima koji proizvode eho **(3)**.

$$\bar{P}_y = 0.98\bar{P}_y + 0.02|\hat{y}|^2, \quad (2)$$

$$\bar{P}_{ref} = 0.98\bar{P}_{ref} + 0.02P_{ref}. \quad (3)$$

Estimacija odnosa ove dve snage se određuje veličinom C_s :

$$C_s = \frac{\sum_{f=0}^{f_s/2} \bar{P}_y(f)}{\sum_{f=0}^{f_s/2} \bar{P}_{ref}(f)}, \quad (4)$$

i ona se koristi za skaliranje snaga zvučničkih signala za potrebe donošenja meke odluke u bloku 408. U ovom bloku se određuje odsustvo bližeg govornika u mikrofonskom signalu na bazi meke odluke definisane relacijom:

$$D_{id} = \alpha_f \log_{10} \frac{\lambda(C_s \bar{P}_{ref}) + \delta}{|e|^2 + \delta}, \quad (5)$$

gde je: α_f - frekvencijski zavisna konstanta kojom se veštački favorizuje dozvola za konvergenciju na višim frekvencijama, gde su snage signala manje, a time i manja mogućnost divergencije NLMS algoritama. Veličina λ je minimalni odnos snage eho signala i bliskog govornika za koji je meka odluka pozitivan broj. U bloku 409 vrši se ograničavanje kontrolnog signala D_{id} , koji se pored NLMS algoritama vodi i u blok DOA-azimut.

Slika 5 prikazuje blok dijagram rešenja za određivanje azimuta 304, odnosno pravca dolaska direktnog zvučnog talasa DOA-azimut od aktivnog govornika. Ulazni signali u ovaj blok su kanalski signali iz AEC bloka s_{AEC1} do s_{AEC5} , a izlazni signal je estimacija dolaznog ugla θ_a . Algoritam se bazira na kroskorelacionoj analizi ulaznih signala s_{AEC1} do s_{AEC5} u bloku 501, na čijem se izlazu dobijaju estimacije četiri kroskorelacione funkcije $G_{1,2}(t,f)$ do $G_{1,5}(t,f)$ rekursivnim usrednjavanjem prema relaciji

$$G_{1,k}(t,f) = \begin{cases} \alpha_+ G_{1,k}(t-1,f) + (1-\alpha_+) X_1(t,f) X_k^*(t,f), & \text{za } |G_{1,k}(t-1,f)| < |X_1(t,f) X_k^*(t,f)| \\ \alpha_- G_{1,k}(t-1,f) + (1-\alpha_-) X_1(t,f) X_k^*(t,f), & \text{za } |G_{1,k}(t-1,f)| \geq |X_1(t,f) X_k^*(t,f)| \end{cases} \quad (6)$$

Konstante α_+ i α_- se biraju tako da ispunjavaju nejednakost $0.5 < \alpha_+ < \alpha_- < 1$ i pod tim uslovom favorizuje se uticaj članova $X_1(t,f) X_k^*(t,f)$ sa najvećim modulom.

U bloku 502 sa oznakom **PHAT** realizuje se generalizovana kroskorelacija u literaturi često označena kao fazna transformacija. Naime, normalizacijom kroskorelacije na svoj moduo gubi se informacija o snazi signala, a ostaje samo informacija o fazi u kojoj je sadržano relativno vremensko kašnjenje signala. Inverznom FFT transformacijom $G_{1,k}(t,f)$ i nalaženjem maksimuma, ocenjuje se relativno vremensko kašnjenje zvučnog talasa između dva mikrofona.

Pošto govorni signal ima formantnu strukturu, zbog čega svi frekvencijski binovi nemaju istu snagu, potrebno je selektovati binove sa najvećom snagom i njih iskoristiti za određivanje kroskorelacione funkcije. U tom cilju se u bloku **503** vrši računanje trenutne snage svakog kanalskog signala i računanje srednje vrednosti snage svih kanala $P(t, f)$. U bloku **504** određuje se težinska funkcija $W(t, f)$ kojom se favorizuju binovi kod kojih postoji rast trenutne snage signala. Razlog izbora ovakvog rešenja je taj što na delu signala sa naglim rastom snage veći je udeo direktnog talasa nego na delu sa padom snage, gde dominiraju refleksije talasa odnosno reverberacija prostorije. U bloku **505** računa se srednja snaga mikrofonskih signala usrednjena po vremenu i po frekvenciji, $\bar{P}(t, f)$. Prvo se vrši usrednjavanje binova po frekvenciji nekauzalnim IIR filtrom prvog reda (nulto fazno kašnjenje se postiže dvostrukim filtriranjem unapred i unazad). Usrednjavanje po vremenu vrši se nelinearnim IIR filtrom prvog reda sa dva koeficijenta usrednjavanja, jedan za rast i drugi za pad snage signala. Ovaj nelinearni filter se opisuje relacijama:

$$\bar{P}(t, f) = \begin{cases} \alpha_{p+} \bar{P}(t, f) + (1 - \alpha_{p+}) P(t, f) \\ \alpha_{p-} \bar{P}(t, f) + (1 - \alpha_{p-}) P(t, f) \end{cases}, \quad 0.8 < \alpha_{p+} < \alpha_{p-} < 1. \quad (7)$$

Veličina $\bar{P}(t, f)$ koristi se za definisanje praga odluke za izdvajanje binova sa najvećom snagom u bloku **506**. Množenjem binarnog izlaza iz bloka **506** i težinskog vektora $W(t, f)$ dobija se filterska funkcija $\bar{W}(t, f)$, kojom se ponderišu binovi fazne transformacije u bloku **502**. Fazno transformisane kroskorelacione funkcije se dodatno filtriraju IIR filtrom u vremenu kako bi se umanjila varijansa estimacije korelacionih funkcija. Ovo se opisuje relacijom:

$$\tilde{G}_{1,k}(t, f) = \alpha_G \tilde{G}_{1,k}(t-1, f) + (1 - \alpha_G) \frac{G_{1,k}(t, f)}{|G_{1,k}(t, f)|}, \quad 0.85 < \alpha_G < 0.95. \quad (8)$$

Pored selekcije binova sa funkcijom $W(t, f)$, primenjuje se i apriorno odbacivanje binova koji se nalaze izvan opsega od interesa. U bloku **507** definisani su opsezi koji apriorno nisu od interesa i oni se odbacuju pre inverzne FFT (FFT^{-1}). U bloku **509** vrši se vremensko usklađivanje kroskorelacionih funkcija, koje se zatim usrednjavaju i na njihovoj srednjoj vrednosti se određuje maksimum u bloku **510**, čija apscisa predstavlja estimaciju vremenskog kašnjenja τ_{est} . U bloku **511** vrši se preračunavanje vremenskog kašnjenja τ_{est} u upadni ugao θ_{est} talasa aktivnog govornika.

Estimacija dolaznog pravca ima smisla kada je bliski govornik aktivan; kada nije aktivan za validnu estimaciju se usvaja estimacija dobijena za vreme poslednje njegove

aktivnosti. U cilju detekcije aktivnosti bliskog govornika koriste se: **a)** informacija iz bloka **513** o srednjoj snazi mikrofonskih signala; **b)** informacija iz detektora dvostruke aktivnosti govornika D_{id} , iz bloka **402**, slika 4; i **c)** informacija s_{BF} iz bloka **303**, **SD-BF** slika 3. Na osnovu ovih informacija u bloku **512** se donosi odluka o aktivnosti bliskog govornika. U slučaju odluke da je estimacija dolaznog pravca validna, da je aktivan bliži govornik, na izlaz DOA bloka **304** se prosleđuje trenutna estimacija dolaznog pravca; u suprotnom se prosleđuje poslednja validna estimacija pravca.

Na slici 6 prikazan je blok dijagram postupka za formiranje superdirektivnog prostornog filtra **303**, slika 3. Zbog problema samoponištavanja korisnog signala koji se javlja kada se adaptivni algoritam za potiskivanje akustičkih smetnji primenjuje u prostoru sa reverberacijom, često se umesto adaptivnog algoritma primenjuje superdirektivni prostorni filter **601** sa fiksnim koeficijentima. Superdirektivni prostorni filter obezbeđuje veći indeks usmerenosti u odnosu na prostorni konvencionalni filter koji sadrži samo kompenzaciju kašnjenja i sumiranje. Opis postupka dobijanja težinskih koeficijenata koji obezbeđuju superdirektivnu karakteristiku filtra su dati u daljem tekstu.

Za prostoru sa reverberacijom se obično usvaja model difuznog polja šuma, što podrazumeva da šum dolazi iz svih pravaca sa približno istim intenzitetom. Za takav model polja šuma pokazuje se da je koherencija između dva mikrofona realan broj jednak

$$\Gamma_{i,j}(f) = \frac{\sin(2\pi f d_{i,j} / c)}{2\pi f d_{i,j} / c}, \quad (9)$$

gde je f učestanost, $d_{i,j}$ je rastojanje mikrofona i i j , a c brzina zvuka. Koherencije parova mikrofona $\Gamma_{i,j}(f)$ formiraju matricu koherencija Γ_d . Koristeći ovako definisanu matricu koherencija Γ_d , koeficijenti superdirektivnog mikrofonskog niza se određuju u bloku **602** prema relaciji:

$$\mathbf{W}_{SD}^H = \frac{\mathbf{C}_\theta^H \Gamma_d^{-1}}{\mathbf{C}_\theta^H \Gamma_d^{-1} \mathbf{C}_\theta}, \quad (10)$$

gde je $\mathbf{C}_\theta$ vektor usmerenja na pravac odabranog govornika definisan azimutom θ . Ovaj vektor se određuje u bloku **603** prema relaciji:

$$\mathbf{C}_\theta^H = \left[1 \exp\left(-j\omega \frac{d \sin(\theta)}{c}\right) \cdots \exp\left(-j\omega \frac{4d \sin(\theta)}{c}\right) \right]. \quad (11)$$

Veličina d je rastojanje dva susedna mikrofona. Na izlazu bloka **303** dobija se estimacija govora s_{BF} aktuelnog govornika na bazi relacije:

$$s_{BF} = \mathbf{W}_{SD}^H \mathbf{S}_{AEC}. \quad (12)$$

Na slici 7 prikazan je blok za potiskivanje šuma **305** sa oznakom **NR**. Signal s_{BF} jeste ulazni signal u blok **305** i on sadrži estimirani govorni signal i rezidualne signale smetnji koji potiču od akustičkog eha, akustičkih smetnji u prostoriji i reverberacije prostorije. Signal s_{BF} se uvodi u blok **701**, označen sa FWF^{-1} , u kome se izvršava IFFT, zatim dopunsko prozorovanje vremenskog oblika segmenta signala u cilju “mekanog” otsecanja krajeva segmenta i na kraju ponovno vraćanje u frekvencijski domen pomoću FFT. Suština ove operacije je sledeća. U procesu prethodnih obrada signala, ekvivalentni vremenski oblik signala se proširuje do granica DFT prozora. Primenom nove operacije Wiener-ovog filtriranja vrši se dodatno proširivanje segmenta i cikličko preklapanje na krajevima segmenta, što stvara impulsne smetnje koje se manifestuju kao ravnomerno “pucketanje”. Primenjeni postupak FWF^{-1} u potpunosti otklanja opisani problem a ne unosi nikakva dodatna izobličenja signala.

U naredna dva bloka **702** i **703** vrši se estimacija šuma na bazi minimuma snage ulaznog signala. Pošto trenutna adaptacija na minimum snage ne daje dobre rezultate, jer DFT koeficijenti na pojedinim blokovima imaju ekstremno nisku snagu koja remeti prethodnu estimaciju snage šuma, estimacija šuma je realizovana u tri bloka obrade. U prvom bloku **702** se vrši spora estimacija snage šuma $\hat{N}_{slow}$, u drugom **703**, brza estimacija snage šuma $\hat{N}_{fast}$, a u trećem **704** se na osnovu procena $\hat{N}_{slow}$ i $\hat{N}_{fast}$ posredstvom nelinearne transformacije vrši procena trenutne snage šuma $\hat{N}$.

Brza i spora procena snage šuma se realizuje istim postupkom rekurzivnog usrednjavanja IIR filtrom prvog reda sa različitim faktorima adaptacije za rast i pad vrednosti izlaza

$$\hat{N}_{slow}(t) = \begin{cases} \alpha_{slow+} \hat{N}_{slow}(t-1) + (1 - \alpha_{slow+}) |S_{SBF}(t)|^2, & \text{za } \hat{N}_{slow}(t-1) < |S_{SBF}(t)|^2 \\ \alpha_{slow-} \hat{N}_{slow}(t-1) + (1 - \alpha_{slow-}) |S_{SBF}(t)|^2, & \text{za } \hat{N}_{slow}(t-1) \geq |S_{SBF}(t)|^2 \end{cases}, \quad (13)$$

$$\hat{N}_{fast}(t) = \begin{cases} \alpha_{fast+} \hat{N}_{fast}(t-1) + (1 - \alpha_{fast+}) |S_{SBF}(t)|^2, & \text{za } \hat{N}_{fast}(t-1) < |S_{SBF}(t)|^2 \\ \alpha_{fast-} \hat{N}_{slow}(t-1) + (1 - \alpha_{fast-}) |S_{SBF}(t)|^2, & \text{za } \hat{N}_{fast}(t-1) \geq |S_{SBF}(t)|^2 \end{cases}, \quad (14)$$

pri čemu između konstanti α_{slow+} , α_{slow-} , α_{fast+} , α_{fast-} postoji relacija:

$$0.2 < \alpha_{fast-} < \alpha_{slow-} < \alpha_{fast+} < \alpha_{slow+} < 1. \quad (15)$$

Brza i spora estimacija šuma se kombinuju u bloku 704, koji je označen kao nelinearni kompresor. Finalna estimacija nivoa šuma se dobija na bazi sledeće relacije:

$$\hat{N} = \begin{cases} \beta \hat{N}_{slow} \left(\frac{\hat{N}_{fast}}{\hat{N}_{slow}} \right)^\alpha & \text{za } \hat{N}_{fast} > \hat{N}_{slow} \\ \beta \hat{N}_{fast} & \text{za } \hat{N}_{fast} \leq \hat{N}_{slow} \end{cases}, \quad (16)$$

gde se parametrom α ($0.25 \leq \alpha < 0.5$) reguliše stepen kompresije dinamike estimacije šuma, a parametrom β definiše se uvećanje estimacije šuma (*overestimation of the noise power*). Smisao nelinearne transformacije je sledeći: u slučaju $\hat{N}_{fast} > \hat{N}_{slow}$ primena samo brze estimacije dala bi prekomerno potiskivanje i govornog signala, zato je uvedena kompresija dinamike estimacije šuma. U slučaju $\hat{N}_{fast} \leq \hat{N}_{slow}$ ne primenjuje se kompresija kako bi estimacija šuma što brže opala. Time se sprečava otsecanje delova fonema na krajevima reči kada zbog brzog pada snage signala visoka vrednost prethodne estimacije šuma sporog estimatora ne može da prati ovu promenu dinamike. Pošto je odnos korisnog govornog signala i šuma znatno nepovoljniji na visokim frekvencijama, definisan je skup parametara α i β za 4 karakteristična opsega frekvencija (0-2000Hz), (2000-2500Hz), (2500-3500Hz) и (3500-5012Hz), prema očekivanom odnosu signal/šum. Ovaj skup parametara je memorisan u bloku 705.

U bloku 706 vrši se Wiener-ovo filtriranje primenom sledeće prenosne funkcije:

$$h_w = \sqrt{\max\left\{\frac{[e_t]^2 - \beta_{oe} \hat{N}}{[e_t]^2}, 0\right\}}, \quad (17)$$

gde konstanta β_{oe} ima funkciju procenjivanja prvobitne procene snage šuma kako bi se ostvario kompromis između što većeg potiskivanja šuma i minimalne degradacije korisnog govornog signala. Prenosna funkcija h_w može imati u vremenskom domenu neprihvatljivo dugačak impulsni odziv, što proizvodi izobličenja na granicama DFT blokova, i zbog toga se vrši “meko” skraćivanje impulsnog odziva primenom gore opisanog postupka FWF⁻¹. Na kraju se vrši u bloku 707 dodatno filtriranje izlaznog estimiranog govornog signala $\hat{s}_t$ kako bi se odbacile spektralne komponente van opsega

govornog signala, koje mogu nastati u prethodnim procesima obrade signala, a koje mogu uticati na rad AGC bloka.

Na slici 8 prikazan je blok za automatsku regulaciju pojačanja (AGC) izlaznog signala sistema, blok 306. Zadatak AGC bloka je: (1) da pojača slabe govorne signale a da oslabi previše jake signale prema unapred zadatoj karakteristici kompresije dinamike signala, (2) da na delovima ulaznog signala gde je prisutan samo eho signala, stacionaran šum ili konkurentni govornik-smetnja, smanji pojačanje kako bi se ove smetnje dovoljno utišale i (3) da utiša delove ulaznog signala gde su jednovremeno prisutni i koristan govorni signal i smetnje, a da pri tome očuva razumljivost govora.

Na ulaz bloka 306 dolazi signal s_{NR} iz bloka NR, slika 3 blok 305, i prolazi kroz kompresor dinamike signala sa adaptivnim nagibom karakteristike kompresije, blok 801. Izlaz iz bloka 801 je signal s_{AGC} koji zatim prolazi kroz blok 307, slika 3, gde se inverznom Fourierovom transformacijom FFT^{-1} konvertuje iz frekvencijskog u vremenski domen i kao konačan signal estimacije govornog signala $\hat{s}$ prenosi ka udaljenom govorniku kroz kanal digitalne televizije.

Kontrola pojačanja govornog signala vrši se u bloku 801 na bazi sledeće relacije:

$$A_{agc} = \left(\frac{P_{nom}}{P_{in} + \alpha P_{nom}} \right)^{(1-SLOPE)/2}, \quad (18)$$

gde su: A_{agc} – pojačanje AGC bloka, P_{nom} – nominalna snaga izlaznog signala, α – konstanta kojom se ograničava maksimalno pojačanje na nivo $A_{agc \max} = \sqrt{1/\alpha}$ (za vrednost $\alpha = 0.001$ maksimalno pojačanje je $A_{agc \max} = 31.6$ dB), $P_{in} = P_d + P_n + P_{echo}$ (P_d – snaga korisnog govornog signala, P_n – snaga difuznog ambijentalnog šuma i P_{echo} – snaga nepotisnutog eho signala), i $SLOPE = f[P_{dp}(t)]$ – veličina koja predstavlja stepen kompresije dinamike signala i složena je funkcija vršne snage korisnog govornog signala. U bloku 802 izračunava se veličina $SLOPE$ na bazi analize trajektorije vršne snage korisnog govornog signala i praćenja njene konveksnosti i trenda rasta.

U bloku 803 izračunava se vršna snaga korisnog govornog signala prema sledećim relacijama:

$$P_{dp}(t) = \begin{cases} P_d(t), & \text{ako je } P_d(t) \geq P_{dp}(t-1) \\ \alpha_d P_{dp}(t-1) + (1 - \alpha_d) P_d(t), & \text{ako je } P_d(t) < P_{dp}(t-1) \end{cases} \quad (19)$$

gde je α_d – konstanta vrednosti blizu 1.

U bloku **804** određuje se estimacija snage nepotisnutog eha prema relaciji:

$$P_{echo} = \alpha_{echo} |\hat{y}|^2, \quad (20)$$

gde je α_{echo} - konstanta potiskivanja eho signala $\hat{y}$ iz bloka **402**, slika 4.

U bloku **805** vrši se estimacija difuznog šuma P_n kao razlika srednje snage ulaznih signala s_{AEC1} do s_{AEC5} u blok **303**, slika 3, i snage izlaznog signala s_{BF} iz bloka **303**.

Neposredna primena relacije za A_{age} za unapred fiksnu veličinu *SLOPE* ne daje dobre rezultate, jer jednako tretira preostale smetnje i koristan signal. Kada su prisutne samo smetnje dolazi do njihovog pojačanja, što nije dobro. Zato je potrebno detektovati i razdvojiti sledeće slučajeve: (a) pauza u korisnom govornom signalu, (b) prisutan rezidualni eho, i (c) prisutan konkurentni govornik ili akustička smetnja. Kada se detektuje bilo koji od ovih slučajeva, promenljiva *SLOPE* se izjednačava sa 1 i tako sprečava pojačanje smetnji.

Pauza u korisnom govornom signalu se razlikuje od govornog signal po stacionarnosti. Govorni signal, ma koliko bio slabog nivoa, nestacionaran je u vremenu, dok je u pauzi prisutan sporopromenjivi ambijentalni šum. Linearni trend snage signala normalizovan na snagu je dobar pokazatelj nestacionarnosti signala. Tome treba dodati i pokazatelj konveksnosti trajektorije koji je negativan na lokalnom maksimumu.

U ovom pronalasku opisan je postupak obrade akustičkih i govornih signala u sistemu slobodne govorne komunikacije koji funkcioniše u punom dupleksu. Ovaj pronalazak se odnosi na slobodnu govornu komunikaciju u sistemu digitalne televizije, ali se isto tako može primeniti i na druge komunikacione sisteme kao što su video-telefonski sistemi, telekonferencijski sistemi, spikerfoni u prostoriji ili kolima, komunikacija čovek-računar putem glasa, i td. Specifičnost rešenja u ovom pronalasku jeste njegova integracija u standardni digitalni TV prijemnik i njegova optimizacija za primenu u prostorijama (akustičkim ambijentima) srednje veličine sa vremenom reverberacije do 600 ms.

Postupci i tehnike obrade akustičkih i govornih signala u ovom pronalasku mogu se generalizovati na N mikrofona u mikrofonskom nizu kod višekanalnog snimanja i na M zvučnika kod višekanalne reprodukcije.

Postupci i tehnike obrade akustičkih i govornih signala u ovom pronalasku se nalaze pod kontrolom većeg broja parametara koji omogućavaju optimizaciju rešenja za različite aplikacije.

Postupci i tehnike obrade akustičkih i govornih signala u ovom pronalasku mogu se implementirati na različite načine. Na primer, ove tehnike mogu biti implementirane u hardveru, softveru ili kombinovano. U hardverskoj implementaciji mogu se koristiti specifična integrisana kola (ASIC), procesori za digitalnu obradu signala (DSP), programabilna logička kola (PLD ili FPGA) i druga elektronska kola projektovana tako da mogu izvršiti opisane funkcije u ovom pronalasku.

Postupci i tehnike obrade akustičkih i govornih signala u ovom pronalasku mogu se implementirati i softverski u celosti ili po modulima koji izvršavaju pojedine funkcije opisane u ovom pronalasku. Programski kodovi mogu biti memorisani u memorijskim jedinicama i izvršavani pomoću procesora kao što su PC, PDA, DSP, itd.

Detalji ovog pronalaska opisani ovde omogućavaju bilo kom stručnjaku u ovoj oblasti da generičke principe ovog pronalaska može implementirati u drugim sistemima za slobodnu govornu komunikaciju čime se ne izlazi iz okvira ovog pronalaska.

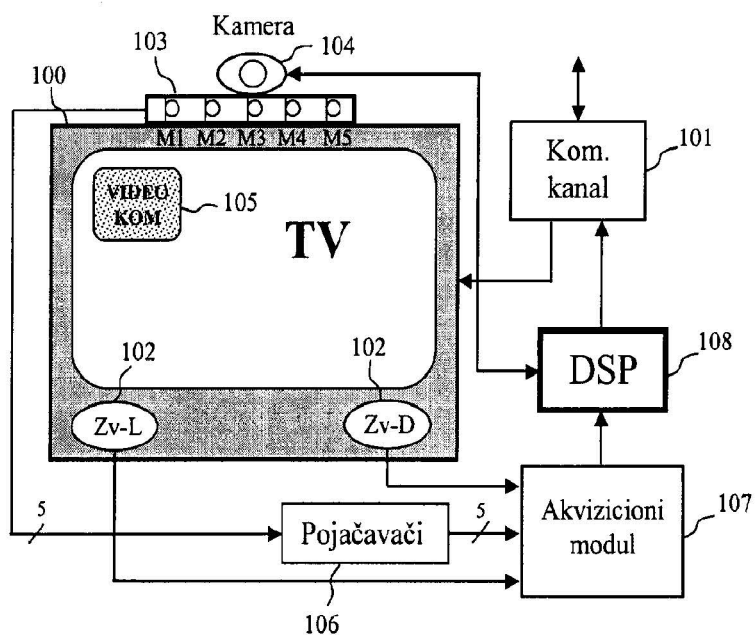


PATENTNI ZAHTEVI

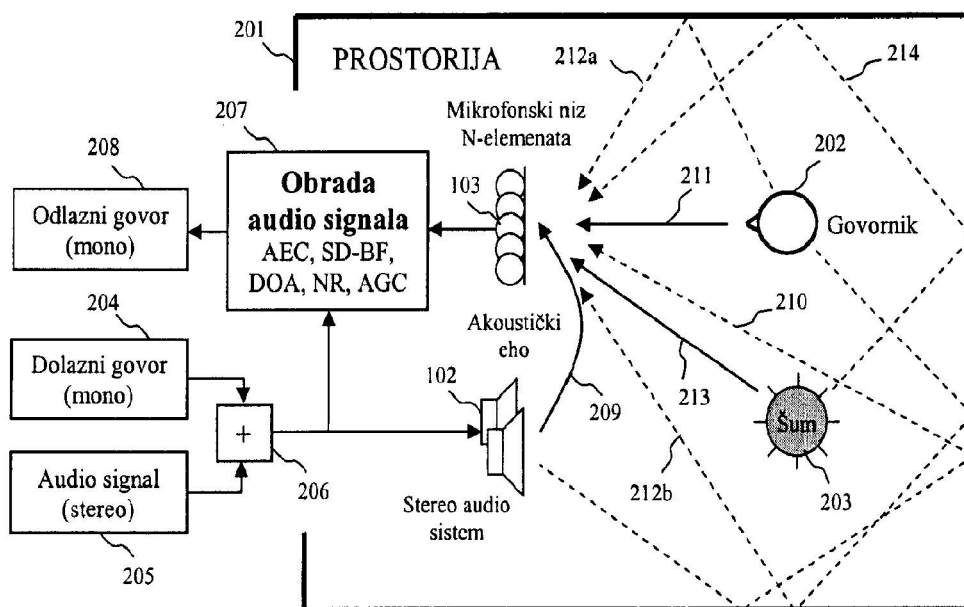
1. Sistem za slobodnu govornu komunikaciju pomoću mikrofonskog niza koji sadrži digitalni TV prijemnik koji omogućava audio i video komunikaciju u punom dupleksu **karakterisan time** što digitalni TV prijemnik (100) ima stereo audio reprodukciju (102) za reprodukovanje stereo TV programa i mono dolaznog govornog signala u videotelefonskoj komunikaciji, koji ima ugrađenu pokretnu video kameru (104) za snimanje govornika u prostoriji i koji na delu svog ekrana reprodukuje sliku sagovornika sa udaljenog kraja (105); koji sadrži mikrofonski sistem (103) ugrađen u TV prijemnik (100) čija je namena snimanje govora govornika na bliskom kraju kao i ostalih ambijentalnih zvukova i čija je namena lociranje govornika u prostoriji i upravljanje video kamerom (104).
2. Sistem prema zahtevu 1 **karakterisan time** što njegov audio predajni deo (207) i (208) omogućava potiskivanje akustičkog eha (209) koji generišu zvučnici TV prijemnika (102), omogućava potiskivanje ambijentalnih smetnji (213) i reverberacije (210), (212) i (214), omogućava lociranje govornika u prostoriji, omogućava adaptivnu kontrolu nivoa signala u predaji i daje koordinate za upravljanje video kamerom.
3. Sistem prema zahtevu 2 **naznačen time** što sadrži mikrofonski niz (103) od više od 2 mikrofona koji obezbeđuju mikrofonske signale za dalju paralelnu obradu, modul za adaptivno potiskivanje akustičkog eha (AEC) (302) koga čini skup adaptivnih filtera, modul za estimaciju dolaznog pravca direktnog zvučnog talasa govornika (DOA) (304) i upravljanje karakteristikom usmerenosti mikrofonskog niza, modul za formiranje karakteristike usmerenosti mikrofonskog niza sa optimiziranim odnosom glavne i bočnih petlji (SB-CBF) (303), modul za adaptivno potiskivanje svih rezidualnih signala smetnji (NR) (305) i modul za automatski kontrolu pojačanja sistema (AGC) (306).
4. Sistem prema zahtevu 3, **naznačen time**, što sadrži skup mikrofona (103) lociranih u horizontalnoj ravni na jednakom međusobnom rastojanju i montiranih na gornjoj ivici digitalnog TV prijemnika (100).
5. Sistem prema zahtevu 4, **naznačen time**, što vrši potiskivanje akustičkog eha (209) koji generišu stereo zvučnici (102) a koji se sastoji od stereo audio TV signala (205) i mono govornog signala koji potiče od udaljenog govornika (204).

6. Sistem prema zahtevu 5, **naznačen time**, što jedinica za potiskivanje eha (302) i jedinica za potiskivanje ambijentalnih smetnji (305) rade i u uslovima malog odnosa signal/šum.
7. Sistem prema bilo kom od prethodnih zahteva **naznačen time**, što omogućava adaptivno lociranje i praćenje govornika u prostoru po azimutu.
8. Sistem prema zahtevu 7, **naznačen time**, što omogućava adaptivno određivanje prostornih koordinata za upravljanje video kamerom.
9. Sistem prema zahtevu 4 **naznačen time**, što njegov mikrofonski niz formira uzanu karakteristiku usmerenosti koja omogućava prostorno filtriranje i separaciju aktuelnog govornika od drugih izvora smetnji u prostoriji.
10. Sistem prema zahtevu 9 **naznačen time**, što njegov mikrofonski niz formira uzanu karakteristiku usmerenosti koja omogućava potiskivanje eha usled refleksija u prostoriji, odnosno signala reverberacije.
11. Sistem prema bilo kom od prethodnih zahteva, **naznačen time**, što putem automatske kontrole pojačanja sistema održava srednji nivo predajnog govornog signala u prihvatljivim granicama normalne dinamike govora bez obzira na udaljenost i položaj govornika u odnosu na mikrofonski niz.
12. Postupak za slobodnu govornu komunikaciju pomoću mikrofonskog niza, **karakterisan time**, što paralelno obrađuje mikrofonske signale iz mikrofonskog niza i time postiže adaptivno potiskivanje akustičkog eha u mikrofonskim signalima, što vrši estimaciju dolaznog pravca direktnog zvučnog talasa bliskog govornika, što formira superdirektivnu karakteristiku usmerenosti mikrofonskog niza i upravlja njenim prostornim položajem po azimutu, što vrši potiskivanje svih signala smetnji koji se nalaze u mikrofonskim signalima i što vrši automatsko održanje nivoa predajnog govornog signala.
13. Postupak prema zahtevu 12, **karakterisan time**, što se kompletna obrada svih audio signala vrši u frekvencijskom domenu.
14. Postupak prema zahtevu 12, **karakterisan time**, što se adaptivno potiskivanje akustičkog eha vrši pojedinačno za svaki mikrofonski signal i da se u potiskivanju obuhvataju oba signala koja dolaze iz stereo zvučnika.
15. Postupak prema zahtevu 14, **karakterisan time**, što se adaptivno potiskivanje akustičkog eha vrši za svaki mikrofonski signal pomoću NLMS algoritama koji su kontrolisani pomoću detektora aktivnosti govora na oba kraja (DTD).

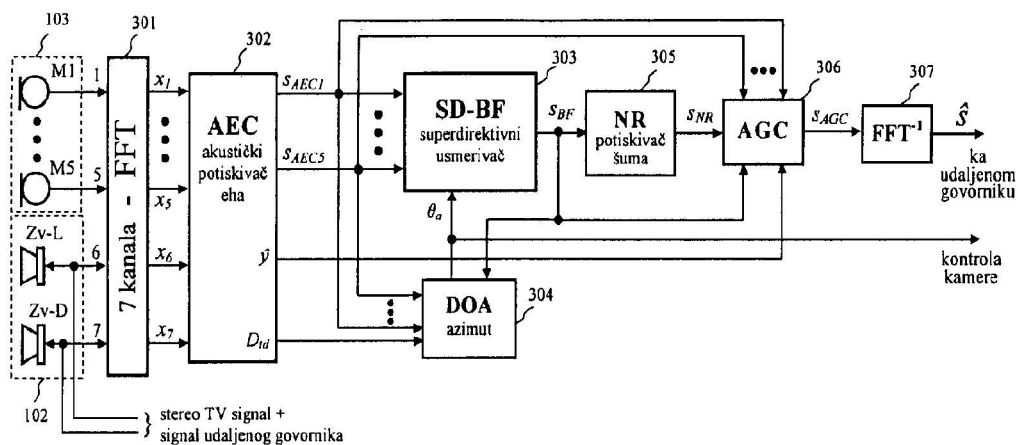
16. Postupak prema zahtevu 14, **karakterisan time**, što se NLMS algoritmi kontrolišu pomoću detektora aktivnosti govora na bližem kraju koji je realizovan u okviru DTD a na bazi RLS adaptivnog algoritma pri specifičnim uslovima kontinualnog prisustva TV audio programskog signala, koji pored govora sadrži i muzički signal.
17. Postupak prema zahtevu 12, **karakterisan time**, što se estimacija dolaznog pravca direktnog zvučnog talasa od aktuelnog govornika vrši na bazi kroskorelacione analize mikrofonskih signala nakon potiskivanja akustičkog eha.
18. Postupak prema zahtevu 17, **karakterisan time**, što se estimacija dolaznog pravca direktnog zvučnog talasa od aktuelnog govornika vrši pod kontrolom VAD detektora za govor na bližem kraju.
19. Postupak prema zahtevu 12, **karakterisan time**, što se karakteristika usmerenosti mikrofonskog niza formira u modulu SB-CBF kao superdirektivna karakteristika na principu ponderisanja i sumiranja mikrofonskih signala nakon potiskivanja akustičkog eha i adaptivnog upravljanja prema azimutu.
20. Postupak prema zahtevu 19, **karakterisan time**, što se koeficijenti superdirektivnog mikrofonskog niza određuju pomoću funkcija koherentnosti parova mikrofonskih signala i vektora usmerenja na pravac odabranog govornika definisan uglom azimuta.
21. Postupak prema zahtevu 12, **karakterisan time**, što se funkcija potiskivanja rezidualnog šuma ostvaruje adaptivnim Wiener-ovim filtrom.
22. Postupak prema zahtevu 21, **karakterisan time**, što je estimacija rezidualnog šuma u potiskivaču šuma optimizirana prema karakteristikama govornog signala i realizovana na bazi nelinearnog kompresora dinamike estimiranog šuma parametarski kontrolisanog i frekvencijski zavisnog.
23. Postupak prema zahtevima 12 do 22 **karakterisan time**, što se modul za automatsku kontrolu pojačanja sistema zasniva na kompresoru dinamike sa adaptivnom nagibom karakteristike kompresije.
24. Postupak prema zahtevu 23, **karakterisan time**, što se kompresor dinamike govornog signala kontroliše pomoću detektora prisutnosti rezidualnog akustičkog eha, detektora pauze u govornom signalu i detektora konkurentnog govornika i akustičkih smetnji.



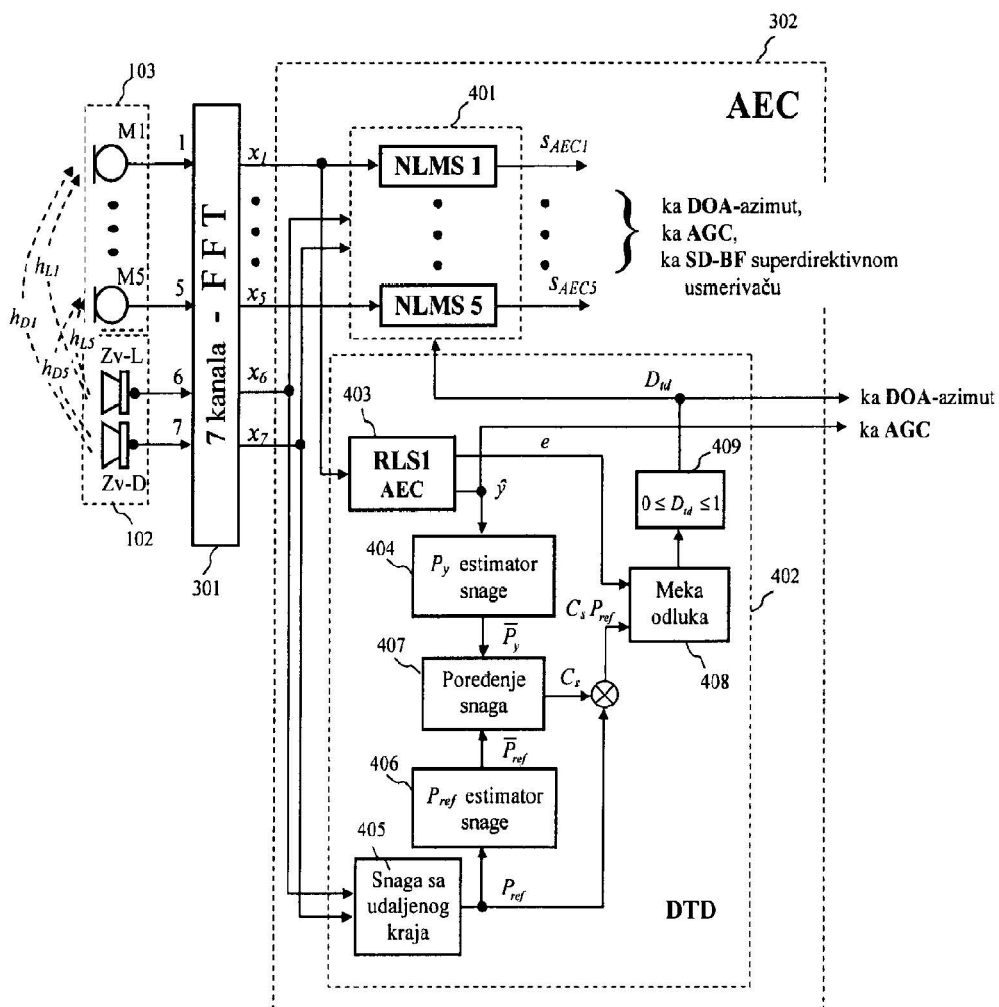
Slika 1.



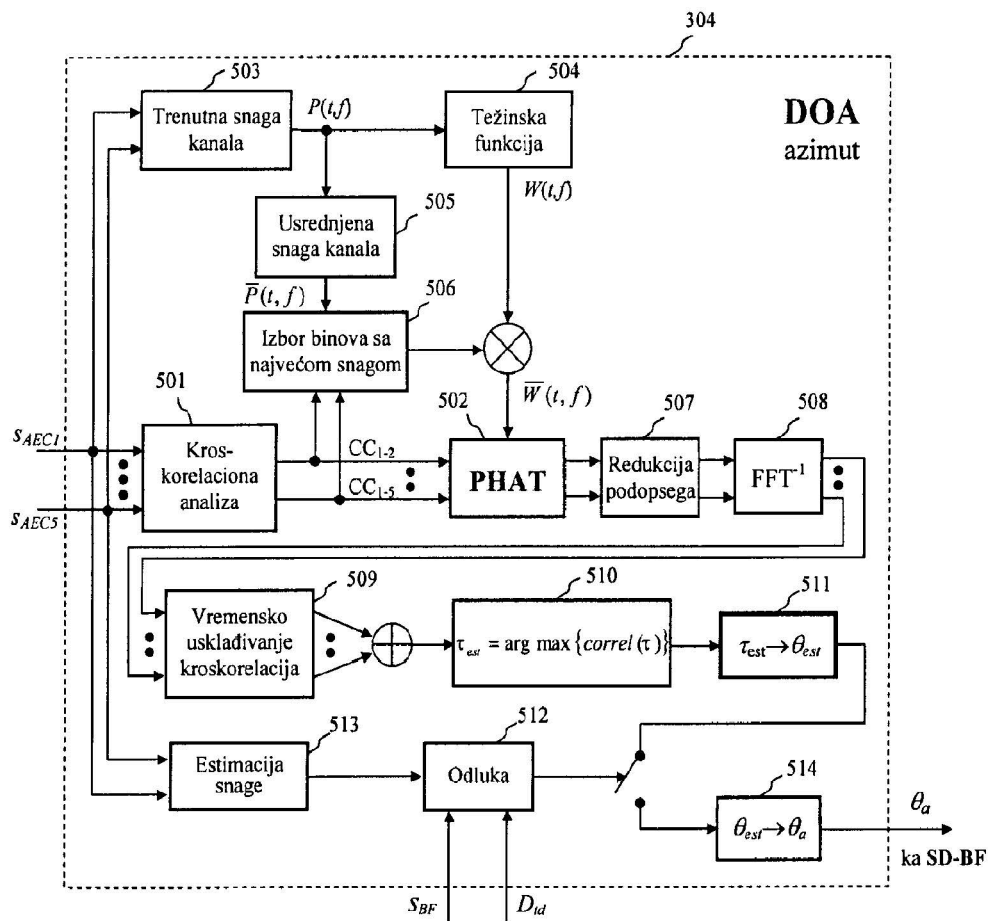
Slika 2.



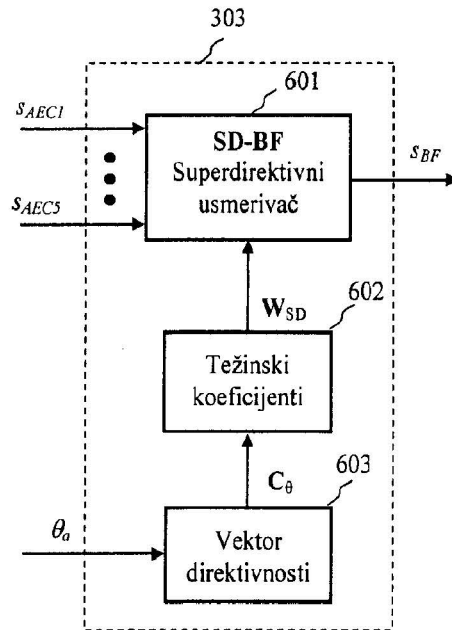
Slika 3.



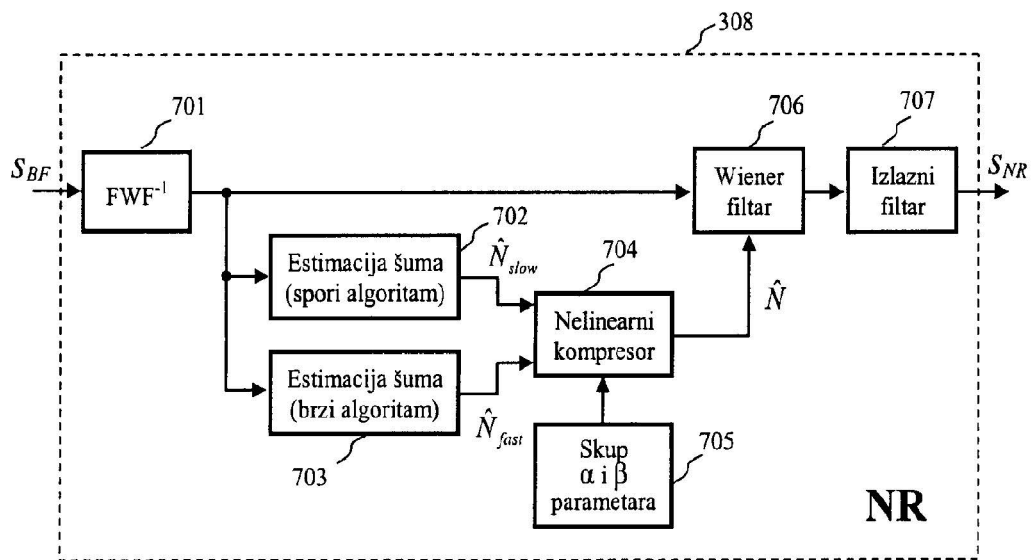
Slika 4.



Slika 5.

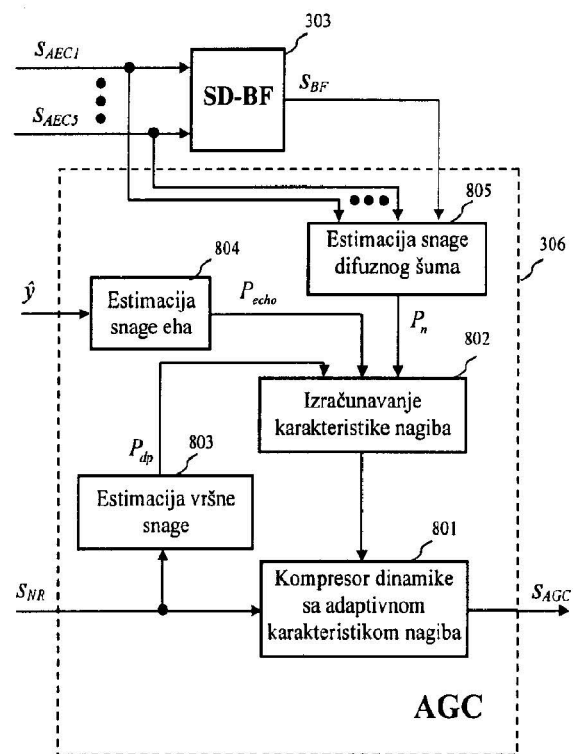


Slika 6.



Slika 7.





Slika 8.