(19) **REPUBLIKA SRBIJA** (12) **Patentni spis** (11) **49859B**

(51) Int. Cl.[7] **H 04 R 29/00**

ZAVOD ZA
INTELEKTUALNU SVOJINU
B E O G R A D

(21) Broj prijave: **P-2006/0642**

(22) Datum podnošenja prijave: **21.11.2006.**

(43) Datum objavljivanja prijave: **04.06.2007.**

(45) Datum objavljivanja patenta: **07.08.2008.**

(30) Međunarodno pravo prvenstva:

(61) Dopunski patent uz osnovni patent broj:

(62) Izdvojen patent iz prvobitne prijave broj:

(73) Nosilac patenta:
MicronasNIT
Fruškogorska 11a
21000 Novi Sad, RS

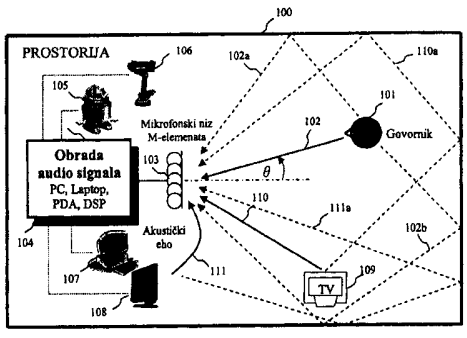
(72) Pronalazači:
Šarić dr Z.;
Jovičić dr S.;
Kovačević dr V.;
Teslić dr N.;
Kukolj dr D.

(74) Zastupnik:

(54) Naziv: **SISTEM I POSTUPAK ZA LOCIRANJE GOVORNIKA POMOĆU MIKROFONSKOG NIZA**

(51) Int. Cl.[7] **H 04 R 29/00**

(57) Apstrakt:
Pronalazak se odnosi na sistem i postupak za lociranje govornika u otvorenom ili zatvorenom prostoru koji je zasnovan na primeni mikrofonskog niza sa proizvoljnim brojem mikrofona i koji može biti primenjen u kontroli i upravljanju robota, video kamere ili procesa koji zahtevaju interaktivnu informaciju o lokaciji govornika, ili u .hands-free" komunikacionim sistemima kao što su telekonferencijski sistemi, video konferencijski sistemi, spikerfoni, itd. Sistem se sastoji od mikrofonskog niza, akvizicionog modula i modula za obradu akustičkih i govornih signala. Postupak se bazira na kroskorelacionoj analizi mikrofonskih signala, na generalizovanoj kroskorelaciji, tj. faznoj transformacij i PHAT, optimizovanoj prema karakteristikama govornog signala i na detektoru aktivnosti govora (VAD) na bazi superdirektivnog usmerivača (SD-BF). Optimizacijom PHAT kroskorelacije postignuta je veća tačnost i preciznost estimacije ugla azimuta θ dok se primenom SD-BF u VAD dobija na pouzdanosti estimiranja u dinamičkim uslovima primene sistema za lociranje govornika.



Slika 1.

RS 49859 B

OBLAST TEHNIKE NA KOJU SE PRONALAZAK ODNOSI

Pronalazak pripada oblasti obrade akustičkog signala, ili preciznije, metodama lociranja govornika primenom mikrofonskog niza u akustičkom ambijentu sa prisutnim šumom i reverberacijom.

TEHNIČKI PROBLEM

Lokalizacija govornika u prostoru je veoma važan tehnički problem u sistemima koji se baziraju na govornoj komunikaciji na relaciji čovek-čovek ili čovek-mašina. On nastaje kao potreba da komunikacija bude što razumljivija uprkos mnogim smetnjama koje se mogu pojaviti u prostoru a koje maskiraju govorni signal. U sistemima kao što su telekonferencijski sistemi ili spikerfoni u prostoriji ili kolima, pored razumljivosti je od primarne važnosti i kvalitet komunikacije. Isti atributi govorne komunikacije su važni i u komunikaciji na primer čoveka i robota, gde robot mora tačno da prepozna govornu komandu. Nešto drugačiji problem se pojavljuje kod upravljanja video kamere, gde kamera treba da se usmeri ka aktuelnom govorniku apstrahujući ostale izvore zvuka u datom ambijentu. Dakle, ne postavlja se problem razumljivosti govora, već separacije govornog signala i ostalih akustičkih signala.

Navedeni primeri govorne komunikacije u prostoru, ili prostoriji, definišu osnovni problem u vidu lokalizacije aktuelnog govornika, odnosno usmeravanje mikrofonskog sistema ka njemu. Mikrofonski sistem može biti usmereni mikrofon ili više mikrofona u odgovarajućem fizičkom rasporedu. Pošto u akustičkom ambijentu pored izvora korisnog signala postoje smetnje veoma različitog porekla, čiji izvori u prostoru mogu biti proizvoljno raspoređeni, mikrofonski sistem mora imati usmerenu karakteristiku osetljivosti i mora se usmeriti ka željenom izvoru signala, tj. aktuelnom govorniku. Drugačije rečeno, mikrofonski sistem mora locirati govornika u horizontalnoj ravni i odrediti ugao azimuta u odnosu na svoje koordinate u prostoru.

Tehnički problem nastaje kada se u ambijentu pojavi veći broj izvora smetnji, kada su ove smetnje nestacionarne, kada se izvori smetnji kreću u prostoru ili kada se

aktuelni govornik kreće u prostoru. Postupak određivanja ugla azimuta govornika mora da reši tri osnovna problema: (1) detekciju govorne aktivnosti aktuelnog govornika, pri tome treba imati u vidu da se u posmatranom prostoru može pojaviti veći broj govornika, (2) separaciju aktuelnog govornika u odnosu na sve ostale izvore smetnji, što podrazumeva potiskivanje signala smetnji a isticanje korisnog signala, i (3) adaptivno praćenje aktuelnog govornika u pokretu, pri čemu se mora uzeti u obzir da i ostali izvori zvuka mogu biti pokretni. Ovaj treći problem se tiče pravilnog usmeravanja sistema (robota, kamere) ka aktuelnom govorniku.

Dodatni problemi se pojavljuju kod lokalizacije govornika u prostoriji sa izraženom reverberacijom. Signali refleksija od zidova ili objekata u prostoriji mogu biti, u zavisnosti od položaja govornika, izvora smetnji i mikrofonskog sistema, znatno jači od direktnog zvučnog talasa aktuelnog govornika.

Iz izloženog se vidi da su tehnički problemi u rešenju lokalizacije govornika u prostoru veoma složeni i da zahtevaju kompleksan 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

Lociranje govornika u uslovima prisustva akustičkih smetnji i reverberacije prostorije predstavlja složen problem. U uslovima kada se spektri korisnog govornog signala preklapaju sa spektrima prisutnih smetnji lociranje govornika se može rešiti na pouzdan način primenom mikrofonskog niza i odgovarajuće obrade signala koja uzima u obzir specifičnosti uslova primene sistema za lociranje. Teorijske osnove u primeni mikrofonskih nizova za lokalizaciju govornika date su u M.S. Brandstein, D.B. Ward (Eds.), *Microphone Arrays: Signal Processing Techniques and Applications*, Springer, Berlin 2001; i u Y. Huang, J. Benesty, *Audio signal processing for next generation multimedia communication systems*, Kluwer Academic Publishers Publ., 2004.

Postoji veliki broj patentiranih rešenja na bazi mikrofonskih nizova kao što su na primer: U.S. objavljena patentna prijava 2003/0051532 A1, prijavljena 15. avgusta 2002., sa naslovom „Robust talker localization in reverberant environment", daje rešenje lokalizacije govornika u reverberantnoj prostoriji na bazi mikrofonskog sistema cirkularne konfiguracije i na bazi energetske detekcije direktnog zvučnog talasa; zatim U.S. patent 6,970,796 B2, prijavljen 1. marta 2004., sa naslovom "System and method

for improving the precision of localization estimates", daje rešenje koje, pored konvencijalnog određivanja DOA sa mikrofonskim nizom, ima sistem za post-procesiranje na bazi statističkog klasterovanja inicijalnih estimacija lokacija i dobijanja finalne estimacije lokacije sa povećanom preciznošću i pouzdanošću; zatim U.S. patent 6,999,593 B2, prijavljena 28. maja 2003., sa naslovom "System and process for robust sound source localization", daje rešenje za lokalizaciju govornika na bazi mikrofonskog niza kombinovanjem težinske kros-korelacije i podešene usmerene karakteristike parova mikrofonskog niza; zatim U.S. objavljena patentna prijava 2005/0080619 A1, prijavljena 13. oktobra 2004, sa naslovom „Method and apparatus for robust speaker localization and automatic camera steering system employing the same", daje rešenje koje lokalizaciju govornika određuje pomoću MUSIC tehnologije.

Generalno, metode estimacije pravca lociranja govornika (izvora zvuka) se mogu podeliti u tri osnovne grupe: metode na bazi superdirektivne karakteristike usmerenosti mikrofonskog niza, metode na bazi kompleksne estimacije spektra mikrofonskih signala i metode na bazi vremenskog kašnjenja zvučnih talasa do mikrofonskog niza TDOA (*Time Delay of Arrival*). Metode iz prve i druge grupe su osetljive na frekvencijske karakteristike svih izvora zvuka u analiziranoj prostoriji i bez a priornog znanja ne pružaju zadovoljavajuću tačnost. U praksi se najčešće koriste TDOA metode (P. Julian et al., A comparative study of sound localization algorithms for energy aware sensor network nodes, *IEEE Trans. Circuits and Systems*, Vol. 51, No. 4, pp. 640-648, Apr. 2004.). U konvencionalnom postupku u prvom koraku vrši se estimacija TDOA za svaki par mikrofona u mikrofonskom nizu. Estimacija se zasniva na kroskorelacionoj analizi koja se u drugom koraku ponderiše težinskom funkcijom PHAT (*Phase Transform*), koja povećava robusnost algoritma procene dolaznih pravaca na prisustvo šuma i reverberacije u prostoriji. Međutim, reverberacija prostorije, prisustvo više izvora zvuka i šuma predstavljaju i dalje veliki problem za ove metode lokalizacije govornika.

IZLAGANJE SUŠTINE PRONALASKA

Predmet ovog pronalaska je sistem i postupak za lociranje govornika pomoću mikrofonskog niza u složenom akustičkom ambijentu koji pored aktuelnog govornika sadrži mnoge signale smetnji kao što su: ambijentalna buka, izvori akustičkih smetnji, reverberacija prostorije i drugi govornici. Kao takav, sistem može naći široku primenu u

sistemima za govornu komunikaciju kao u sistemima za kontrolu i upravljanje putem glasa.

Sistem, koji je predmet pronalaska, sadrži M mikrofona raspoređenih u linijskoj strukturi i na jednakim rastojanjima, blok za predobradu i digitalizaciju mikrofonskih signala i blok za digitalnu obradu mikrofonskih signala. Sistem se postavlja u horizontalnu ravan i određuje ugao azimuta govornika u odnosu na simetralu sistema. Sistem može biti povezan na konferencijski sistem, ili biti deo njega, ili na sistem upravljanja ili kontrole, kao što su robot ili video kamera.

Suština pronalaska jeste u specifičnoj obradi govornog signala koji se snima u akustičkom ambijentu prostorije u kojoj se nalazi sistem i govornik. 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 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 aktuelnog 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).

Mikrofonski signali iz mikrofonskog niza se obrađuju u digitalnoj formi u frekvencijskom domenu. Ovaj domen omogućava određene prednosti u pogledu brzine obrade i broja računskih operacija, što je veoma važno za realizaciju sistema u realnom vremenu.

Specifičan aspekt pronalaska se nalazi u optimizaciji kroskorelacione analize mikrofonskih signala kroz dva aspekta: prvo, generalizacijom kroskorelacije koja se u literaturi označava kao fazna transformacija PHAT (*Phase Transform*), a koja podrazumeva normalizaciju kroskorelacije na svoj moduo kada se gubi informacija o snazi signala, a ostaje samo informacija o fazi u kojoj je sadržano relativno vremensko kašnjenje signala i drugo, ponderisanjem PHAT transformacije filterskom funkcijom $\overline{W}(n)$ koja sadrži osnovne prozodijske karakteristike govornog signala, pre svega energetsku dinamiku formantnih struktura vokala.

Sledeća specifičnost pronalaska jeste određivanje filterske funkcije $\overline{W}(n)$ na bazi analize mikrofonskih signala u tri domena: energetskom, frekvencijskom i vremenskom. Cilj ove analize je da se generalizovana kroskorelaciona analiza odvija

pod kontrolom prozodijskih karakteristika govornog signala, što predstavlja na specifičan način separaciju govornog signala u odnosu na ostale signale ambijentalnih smetnji i reverberaciju, i što u krajnjem slučaju daje pouzdaniju estimaciju lokacije govornika.

Specifičnost pronalaska jeste i realizacija detektora aktivnosti govora (VAD) na bazi superdirektivnog usmerivača (SD-BF), koji obezbeđuje veći indeks usmerenosti mikrofonskog niza i time efikasnije prostorno filtriranje govornog signala u odnosu na ambijentalne smetnje.

Inventivnost u ovom pronalasku se nalazi u načinu realizacije 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, posebno ako se ima u vidu realizacija u spektralnom i multidimenzionalnom domenu (multimikrofonski sistem).

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 ambijentalne uslove primene sistema za lociranje govornika pomoću mikrofonskog niza.

Slika 2 – prikazuje osnovni blok dijagram sistema za lociranje govornika.

Slika 3 – prikazuje blok dijagram podsistema za estimaciju filterske funkcije $\overline{W}(n)$.

Slika 4 – prikazuje blok dijagram podsistema za estimaciju ugla azimuta $\hat{\theta}$.

Slika 5 – prikazuje blok dijagram podsistema VAD za detekciju aktivnosti govora aktuelnog govornika.

DETALJAN OPIS PRONALASKA

Ovaj pronalazak opisuje sistem i postupak za lokalizaciju govornika pomoću mikrofonskog niza u akustičkom ambijentu kakav je prostorija, sa prisutnim stacionarnim i/ili nestacionarnim smetnjama.

Slika 1 prikazuje ambijentalne uslove u kojima se sistem, koji je predmet ovog pronalaska, može naći. Naime, u prostoriji **100** nalazi se aktuelni govornik **101** u horizontalnoj ravni na pravcu **102** pod uglom θ u odnosu na simetralu mikrofonskog niza **103**. Mikrofonski niz sadrži M mikrofona koji snimljene signale prosleđuju u blok **104** gde se vrši obrada signala u cilju određivanja estimacije ugla azimuta $\hat{\theta}$. Informacija o estimiranom uglu azimuta može da se koristi, na primer za kontrolu robota **105** ili video kamere **106**, ili za komunikacione potrebe kao što su govorna komunikacija preko interneta **107** ili preko telekonferencijskog sistema **108**. U drugom slučaju uglom azimuta upravlja se karakteristikom usmerenosti mikrofonskog niza **103**, koja se usmerava prema aktuelnom govorniku.

Osnovni problem u estimaciji ugla azimuta θ čine smetnje u prostoriji koje direktno utiču na tačnost i preciznost estimacije. Osnovni izvor smetnje može biti izvor šuma, govora, muzike, itd., **109**, sa direktnim zvučnim talasom **110**, ali i reflektovanim zvučnim talasima o zidove prostorije, kao što je talas **110a**. Naravno, i aktuelni govornik jeste izvor reflektovanih talasa, **102a** i **102b**, koji predstavljaju smetnju. Ako se mikrofonski niz **103** koristi za komunikacione potrebe, slučajevi **107** i **108**, kod tzv. „hands-free" komunikacija, tada se pojavljuje veoma ozbiljna smetnja u vidu akustičkog eha **111**, koja ima svoje akustičke refleksije **111a**.

Prema tome, tačnost i preciznost određivanja ugla azimuta u velikoj meri zavisi od ambijentalnih uslova u kojima se sistem, koji je predmet ovog patenta, koristi. Dodatni problem se pojavljuje ukoliko se aktuelni govornik ili izvori smetnji kreću u prostoriji, čime se postavlja zahtev adaptivnog praćenja pozicije aktuelnog govornika.

Na slici 2 prikazana je blok šema sistema za lokalizaciju govornika pomoću mikrofonskog niza. Signali iz mikrofona x_1 do x_M mikrofonskog niza **103** ulaze u blok **201** u kome se vrši njihova predobrada, odnosno pojačanje, filtriranje, digitalizacija i konverzija u frekvencijski domen pomoću diskretne Fourierove transformacije (DFT). Predobrada se vrši na nivou segmenata dužine N odmeraka, sa preklapanjem 50% i sa primenjenim Hammingovim prozorom i FFT reda N.

Izlaz bloka **201** jesu Fourierove transformacije X_1 do X_M. Na ovim signalima vrši se kroskorelaciona analiza prvog mikrofona sa svim ostalim mikrofonima. Na izlazu bloka **201** dobijaju se estimacije kroskorelacije između signal provog i svih ostalih mikrofona, $G_{1,2}(n)$ do $G_{1,M}(n)$ rekurzivnim usrednjavanjem prema relaciji:

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

$$k=2,\ldots,\mathrm{M}$$

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 većim modulom. Signali $G_{1,2}(n)$ do $G_{1,\mathrm{M}}(n)$ ulaze u blokove **203** i **205**.

U bloku **203** 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.

$$G_{1,k\,Phat}(n)=\sum_{l=o}^{N-1}\frac{G_{1,k}(n,l)}{\left|G_{1,k}(n,l)\right|}e^{j\frac{2\pi nl}{N}} \quad (2)$$

U obradi generalizovanih kroskorelacionih funkcija $G_{1,2\,Phat}$ učestvuje filterska funkcija $\overline{W}(n)$ koja se generiše u bloku **204**. Funkcija $\overline{W}(n)$ se dobija obradom mikrofonskih signala X_1 do X_M, koja će kasnije biti detaljnije opisana, a čiji je cilj da osnovne prozodijske karakteristike govornog signala, pre svega energetsku dinamiku formantnih struktura vokala, iskoristi za pouzdaniju ocenu ugla azimuta, odnosno lokaciju govornika u prostoriji.

U bloku **205** vrši se određivanje estimacije ugla azimuta $\hat{\theta}$ na bazi maksimuma generalizovanih kroskorelacionih funkcija. Validnost date estimacije kontroliše blok **206**, sa oznakom **VAD**, koji vrši detekciju aktivnosti aktuelnog govornika, i kada je govornik aktivan validna je tekuća estimacija ugla azimuta, u suprotnom usvaja se estimacija dobijena za vreme poslednje njegove aktivnosti.

Na slici 3 prikazana je blok šema podsistema za određivanje filterske funkcije $\overline{W}(n)$. 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 **301** vrši računanje srednje snage mikrofonskih signala X_1 do X_M po svakom DFT binu unutar bloka n, tj. trenutne snage kanala prema relaciji:

$$P(n)=\frac{1}{M}\sum_{k=1}^{M}\left|X\right|_k^2(n). \quad (3)$$

U bloku **302** određuje se težinska funkcija *W(n)* kojom se favorizuju binovi kod kojih postoji rast trenutne snage signala. Razlog izbora ovakvog rešenja je taj što je na delu signala sa naglim rastom snage veći udeo direktnog talasa nego na delu sa padom snage, gde dominiraju refleksije talasa, odnosno reverberacija prostorije. Ovaj pristup se realizuje relacijom:

$$W(n) = \max\left\{\frac{P(n) - P(n-1)}{P(n)}, 0\right\}. \tag{4}$$

U bloku **303** vrši se dalja obrada kanalskih trenutnih snaga glačanjem (*smoothing*, engl.) snaga *P(n)* po frekvenciji, snaga $\widetilde{P}(n)$, a zatim usrednjavanjem po vremenu, snaga $\overline{P}(n)$. Glačanje snage $P(n)$ vrši se nekauzalnim IIR filtrom prvog reda (nulto fazno kašnjenje se postiže dvostrukim filtriranjem unapred i unazad), tako da se dobija snaga $\widetilde{P}(n)$, dok se usrednjavanje ove snage po vremenu vrši nelinearnim IIR filtrom prvog reda sa dva koeficijenta usrednjavanja, jedan za rast i drugi za pad snage signala. Ovaj nelinearni filtar se opisuje relacijama:

$$\overline{P}(n) = \begin{cases} \alpha_{p+}\overline{P}(n-1) + (1-\alpha_{p+})\widetilde{P}(n), & za\ \ \widetilde{P}(n) > \overline{P}(n-1) \\ \alpha_{p-}\overline{P}(n-1) + (1-\alpha_{p-})\widetilde{P}(n), & za\ \ \widetilde{P}(n) \le \overline{P}(n-1) \end{cases}, \tag{5}$$

$$0.8 < \alpha_{p+} < \alpha_{p-} < 1$$

Veličina $\overline{P}(n)$ koristi se za definisanje praga odluke za izdvajanje binova sa najvećom snagom u bloku **304**. Postupak se sastoji u poređenju veličine $\overline{P}(n)$ i kroskorelacionih funkcija $G_{1,2}(n)$ do $G_{1,M}(n)$ sa binarnom odlukom na izlazu za svaki bin. To znači da se na izlazu bloka **304** dobija M-1 binarnih nizova dužine N.

Množenjem binarnog izlaza iz bloka **304** i težinske funkcije *W(n)* iz bloka **302** dobija se filterska funkcija $\overline{W}(n)$ na uzlazu bloka **305**, kojom se ponderišu binovi fazne transformacije $G_{1,k\,Phat}(n)$ u bloku **203**, slika 2. Fazno transformisane kroskorelacione funkcije se dodatno filtriraju IIR filtrom u vremenu kako bi se umanjila varijansa estimacije korelacionih funkcija. Ovo se opisuje relacijom:

$$\widetilde{G}_{1,k\,Phat}(n) = \alpha_G \widetilde{G}_{1,k\,Phat}(n-1) + (1-\alpha_G)\,\overline{W}(n)\frac{G_{1,k}(n)}{\left|G_{1,k}(n)\right|}, \quad 0.85 < \alpha_G < 0.95. \tag{6}$$

Na slici 4 prikazana je detaljna blok šema bloka **205** sa slike 2, u kome se vrši određivanje estimacije ugla azimuta $\hat{\theta}$. Fazno transformisane kroskorelacione funkcije

$\widetilde{G}_{1,k\,Phat}$ se u bloku **401** pomoću inverzne Fourierove transformacije (IFFT) transformišu iz frekvencijskog u vremensi domen u kroskorelacije $R_{1,2}(\tau)$ do $R_{1,M}(\tau)$. Pre IFFT transformacije primenjuje se u bloku **402** apriorno odbacivanje binova koji se nalaze izvan opsega od interesa. Kriterijum ovog odbacivanja je izbor opsega frekvencija za koji je snaga govornog signala dovoljno velika a da za najveću frekvenciju opsega ne dolazi do alijasinga u prostornom domenu.

U bloku **403** vrši se vremensko usklađivanje kroskorelacionih funkcija $R_{1,2}(\tau)$ do $R_{1,M}(\tau)$ primenom odgovarajućih faktora interpolacije, koje se zatim usrednjavaju i na njihovoj srednjoj vrednosti $R_{sr}(\tau)$ se određuje maksimum u bloku **404**, čija apscisa predstavlja estimaciju vremenskog kašnjenja $\hat{\tau}$.

$$\hat{\tau} = \arg\max_{\tau} R_{sr}(\tau) \tag{7}$$

U bloku **405** vrši se preračunavanje vremenskog kašnjenja $\hat{\tau}$ u upadni ugao $\hat{\theta}_R$ direktnog talasa aktivnog govornika. Estimacija dolaznog pravca ima smisla kada je govornik aktivan; kada nije aktivan za validnu estimaciju se usvaja estimacija dobijena za vreme poslednje njegove aktivnosti. U tu svrhu u bloku **406** se pod kontrolom signala VAD definitivno određuje validnost estimacije $\hat{\theta}_R$, tako da se na izlazu dobija konačna vrednost estimiranog ugla azimuta $\hat{\theta}$.

U cilju detekcije aktivnosti govornika koriste se: a) informacija iz bloka **301** o srednjoj snazi mikrofonskih signala *P(n),* slika 3, i b) informacija s_{BF} iz bloka **501**, blok **SD-BF** superdirektivni usmerivač, slika 5. Na osnovu ovih informacija u bloku **502** se donosi odluka o aktivnosti bliskog govornika.

Formiranje superdirektivnog prostornog filtra vrši se u bloku **501**. On obezbeđuje veći indeks usmerenosti u odnosu na prostorni konvencionalni filter koji sadrži samo kompenzaciju kašnjenja i sumiranje.

Za prostoriju 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:

$$\Gamma_{i,j}(f) = \frac{\sin(2\pi f d_{i,j}/c)}{2\pi f d_{i,j}/c}, \tag{8}$$

gde je f učestanost, d_{ij} je rastojanje mikrofona i i j, a c brzina zvuka. Koherencije parova mikrofona $\Gamma_{i,j}(f)$ formiraju matricu koherencija $\mathbf{\Gamma}_d$. Koristeći ovako definisanu matricu koherencija $\mathbf{\Gamma}_d$, koeficijenti superdirektivnog mikrofonskog niza se odredjuju u bloku **504** prema relaciji:

$$\mathbf{W}_{SD}^{H} = \frac{\mathbf{C}_{\theta}^{H}\mathbf{\Gamma}_{d}^{-1}}{\mathbf{C}_{\theta}^{H}\mathbf{\Gamma}_{d}^{-1}\mathbf{C}_{\theta}}, \tag{9}$$

gde je $\mathbf{C}_\theta$ vektor usmerenja na pravac odabranog govornika definisan estimiranim uglom azimuta $\hat{\theta}$. Ovaj vektor se određuje u bloku **503** prema relaciji:

$$\mathbf{C}_{\theta}^{H} = \left[1 \ \exp\left(-j2\pi f \frac{d\sin(\hat{\theta})}{c}\right) \cdots \exp\left(-j2\pi f \frac{4d\sin(\hat{\theta})}{c}\right)\right]. \tag{10}$$

Veličina d je rastojanje dva susedna mikrofona.

Na izlazu bloka **501** dobija se estimacija govora s_{BF} aktuelnog govornika na bazi relacije:

$$s_{BF} = \mathbf{W}_{SD}^{H} X \tag{11}$$

Prema tome, u blok **502** dolaze dve informacije: srednja snaga mikrofonskih signala $P(n)$, koja pored aktuelnog govornika sadrži i sve signale smetnji u prostoriji, i signal estimacije govora s_{BF} aktuelnog govornika na pravcu estimiranog ugla azimuta $\hat{\theta}$. U bloku **502** se vrši binarna odluka o aktivnosti aktuelnog govornika na bazi komparativne analize prispelih informacija i binarni signal VAD odlučuje o izlaznoj vrednosti ugla azimuta $\hat{\theta}$, odnosno na izlaz sistema se prosleđuje trenutna estimacija dolaznog pravca ako je aktivan aktuelni govornik, u suprotnom se prosleđuje poslednja validna estimacija pravca.

U ovom pronalasku opisan je postupak obrade akustičkih i govornih signala u cilju lokacije govornika u prostoru u odnosu na sistem za lokaciju govornika koncipiranog na bazi mikrofonskog niza. Opisanim sistemom se govornik može locirati u zatvorenom ili otvorenom prostoru, a sistem se može primeniti u kontroli i upravljanju robota, video kamere ili procesa koji zahtevaju interaktivnu informaciju o lokaciji govornika, ili u „hands-free" komunikacionim sistemima kao što su telekonferencijski sistemi, video konferencijski sistemi, spikerfoni, itd.

Postupci i tehnike obrade akustičkih i govornih signala u ovom pronalasku su nezavisne od broja mikrofona u nizu a nalaze se 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, tako da se programski kodovi mogu memoristi u memorijskim jedinicama i izvršavati 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 čime se ne izlazi iz okvira ovog pronalaska.

PATENTNI ZAHTEVI

1. Sistem za lociranje govornika pomoću mikrofonskog niza **karakterisan time,** što sadrži:

   mikrofonski niz od M mikrofona u odnosu na čiju simetralu se određuje ugao azimuta, odnosno položaj govornika u horizontalnoj ravni;

   blok za predprocesiranje mikrofonskih signala i konverziju u digitalnu formu i frekvencijski domen;

   blok za kroskorelacionu analizu mikrofonskih signala i njenu optimizaciju na bazi fazne transformacije;

   blok za određivanje filterske funkcije na bazi prozodijskih karakteristika govornog signala, pomoću koje se vrši optimizacija kroskorelacione PHAT analize;

   blok za detekciju aktivnosti govora (VAD) zasnovan na superdirektivnom usmerivaču (SD-BF) koji obezbeđuje prostorno filtriranje govornika;

   blok za estimaciju ugla azimuta na bazi maksimuma interpoliranih kroskorelacionih funkcija.

2. Sistem prema zahtevu 1 **karakterisan time,** što sadrži blokove koji vrše detekciju govorne aktivnosti aktuelnog govornika, koji vrše separaciju aktuelnog govornika u odnosu na sve ostale izvore smetnji i koji vrše adaptivno praćenje aktuelnog govornika u pokretu.

3. Sistem prema zahtevu 1 **karakterisan time,** što sadrži mikrofonski niz od M mikrofona i što broj mikrofona u nizu nije ograničavajući faktor.

4. Sistem prema zahtevu 2 **karakterisan time,** što se mikrofonski niz nalazi u horizontalnoj ravni i što se lociranje govornika određuje pomoću ugla azimuta u odnosu na simetralu mikrofonskog niza.

5. Sistem prema zahtevu 1 **karakterisan time,** što se obrada signala odvija u frekvencijskom domenu i što se ista može realizovati u realnom vremenu.

6. Sistem prema bilo kom od prethodnih zahteva **karakterisan time,** što sadrži blok kroskorelacione analize mikrofonskih signala koji određuje vremensko kašnjenje zvučnih talasa od izvora zvuka do mikrofonskog niza (TDOA) i koji vrši njenu optimizaciju kroskorelacione analize na bazi fazne transformacije (PHAT).

7. Sistem prema zahtevu 6 **karakterisan time,** što sadrži blok za određivanje filterske funkcije na bazi prozodijskih karakteristika govornog signala koji omogućava optimizaciju kroskorelacione PHAT analize prilagođenu karakteristikama govornog signala.
8. Sistem prema zahtevima 1 do 5 **karakterisan time,** što sadrži blok za detekciju aktivnosti govora (VAD) u uslovima ambijentalnih smetnji i reverberacije, koji odlučuje o konačnoj vrednosti ugla azimuta.
9. Sistem prema zahtevu 8 **karakterisan time,** što osnovu bloka za detekciju aktivnosti govora (VAD) čini superdirektivi usmerivač (SD-BF) koji obezbeđuje separaciju aktuelnog govornika od ostalih izvora zvuka u prostoriji na bazi prostornog filtriranja.
10. Sistem prema zahtevima 6 do 7 **karakterisan time,** što sadrži blok za estimaciju ugla azimuta na bazi maksimuma interpoliranih i usrednjenih M-1 kroskorelacionih funkcija.
11. Sistem prema bilo kom od prethodnih zahteva **karakterisan time,** što se može primeniti za kontrolu uređaja, sistema ili procesa putem glasa.
12. Sistem prema bilo kom od prethodnih zahteva **karakterisan time,** što se može primeniti u „hands-free" komunikacionim sistemima za slobodnu govornu komunikaciju u cilju poboljšanja kvaliteta i razumljivosti komunikacije u akustičkom ambijentu.
13. Postupak za lociranje govornika pomoću mikrofonskog niza **karakterisan time,** što sadrži:

kroskorelacionu analizu koja vrši analizu vremenskog kašnjenja zvučnih talasa od izvora zvuka do mikrofonskog niza;

generalizaciju kroskorelacione analize, odnosno njenu faznu transformaciju (PHAT), uz primenu adaptivnog ponderisanja filterskom funkcijom $\overline{W}(n)$;

adaptivno određivanje filterske funkcije $\overline{W}(n)$ na bazi prozodijskih karakteristika govornog signala;

adaptivnu detekciju aktivnosti govornika (VAD) na bazi superdirektivnog usmerivača (SD-BF);

interpolaciju kroskorelacionih funkcija i određivanje estimacije ugla azimuta.

14. Postupak prema zahtevu 13 **karakterisan time,** što se kroskorelacija vrši između prvog mikrofonskog signala i svih ostalih mikrofonskih signala, tako da se izvršava M-1 kroskorelacija.

15. Postupak prema zahtevu 14 **karakterisan time,** što se generalizacija kroskorelacione analize vrši normalizacijom kroskorelacije na svoj moduo pri čemu se gubi informacija o snazi signala, a ostaje samo informacija o fazi u kojoj je sadržano relativno vremensko kašnjenje između analiziranih signala.

16. Postupak prema zahtevu 15 **karakterisan time,** što se fazno transformisane kroskorelacione funkcije dodatno adaptivno filtriraju IIR filtrom u vremenu kako bi se umanjile varijanse estimacije korelacionih funkcija.

17. Postupak prema zahtevu 16 **karakterisan time,** što se dodatno filtriranje fazno transformisanih kroskorelacionih funkcija vrši adaptivnim ponderisanjem binova fazne transformacije filterskom funkcijom $\overline{W}(n)$.

18. Postupak prema zahtevu 13 **karakterisan time,** što se filterska funkcija $\overline{W}(n)$ određuje na bazi prozodijskih karakteristika govornog signala detektovanih u trenutnoj snazi mikrofonskih signala.

19. Postupak prema zahtevu 18 **karakterisan time,** što se filterskom funkcijom $\overline{W}(n)$ selektuju binovi sa najvećom snagom i oni koriste za određivanje kroskorelacionih funkcija.

20. Postupak prema zahtevu 18 i 19 **karakterisan time,** što se u određivanju filterske funkcije $\overline{W}(n)$ izračunavaju trajektorije snaga mikrofonskih signala usrednjavanjem po frekvenciji i po vremenu.

21. Postupak prema zahtevu 18 do 20 **karakterisan time,** što se u određivanju filterske funkcije $\overline{W}(n)$ favorizuju binovi kod kojih postoji rast trenutne snage signala, iz razloga što je na delu signala sa naglim rastom snage veći udeo direktnog talasa nego na delu sa padom snage, gde dominiraju refleksije talasa, odnosno reverberacija prostorije.

22. Postupak prema zahtevu 13 **karakterisan time,** što se u detektoru aktivnosti govora (VAD) donosi odluka o aktivnosti bliskog govornika.

23. Postupak prema zahtevu 13 i 22 **karakterisan time,** što se VAD bazira na superdirektivnom usmerivaču (SD-BF) koji obradom mikrofonskih signala obezbeđuje usmerenu karakteristiku osetljivosti mikrofonskog niza.

24. Postupak prema zahtevu 23 **karakterisan time,** što superdirektivni usmerivač vrši prostorno filtriranje kojim ističe signal aktuelnog govornika i potiskuje signale ambijentalnih smetnji.

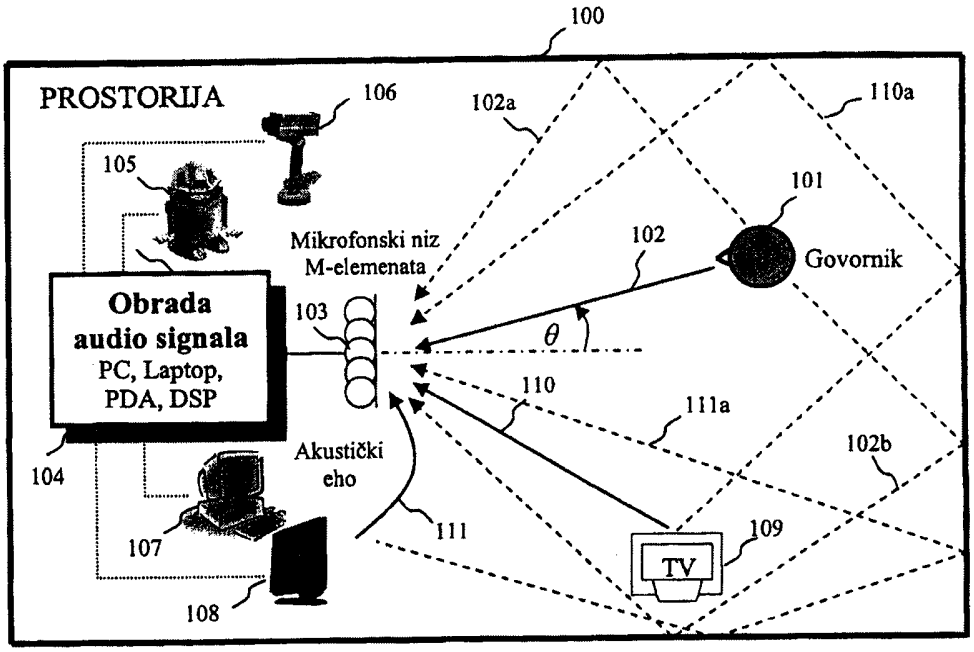
25. Postupak prema zahtevima 23 i 24 **karakterisan time,** što se karakteristikom usmerenosti superdirektivnog usmerivača (SD-BF) upravlja estimiranim uglom azimuta $\hat{\theta}$.

26. Postupak prema zahtevu 13 **karakterisan time,** što se estimacija ugla azimuta $\hat{\theta}$ određuje na bazi maksimuma usklađenih i usrednjenih M-1 kroskorelacionih funkcija.

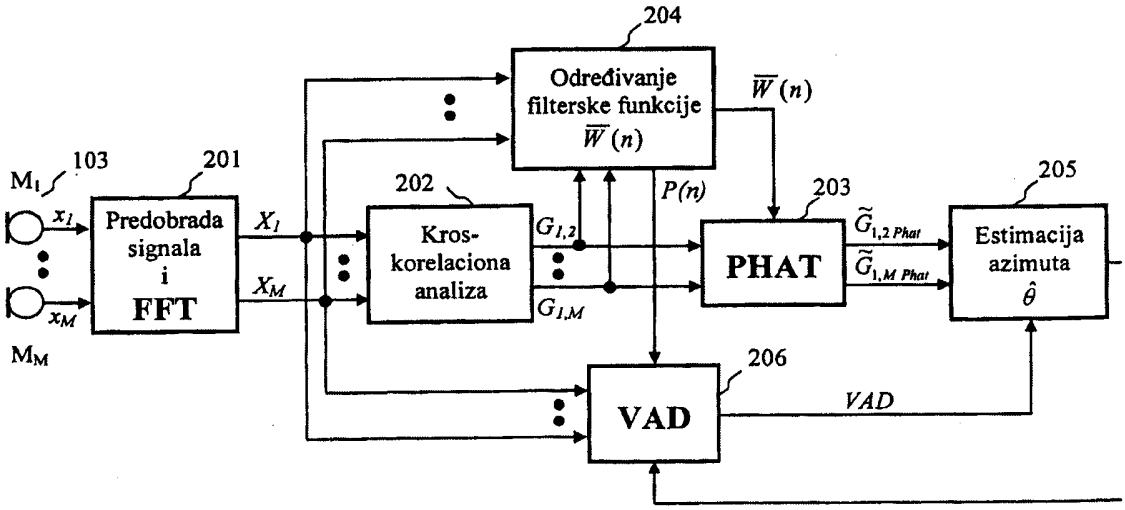
27. Postupak prema zahtevu 26 **karakterisan time,** što se usklađivanje kroskorelacionih funkcija vrši postupkom interpolacije.

28. Postupak prema zahtevima 13, 26 i 27 **karakterisan time,** što se estimacija ugla azimuta $\hat{\theta}$ dobija preračunavanjem vremenskog kašnjenja $\hat{\tau}$ na kome se nalazi maksimum usklađenih i usrednjenih M-1 kroskorelacionih funkcija u upadni ugao θ direktnog talasa aktuelnog govornika.

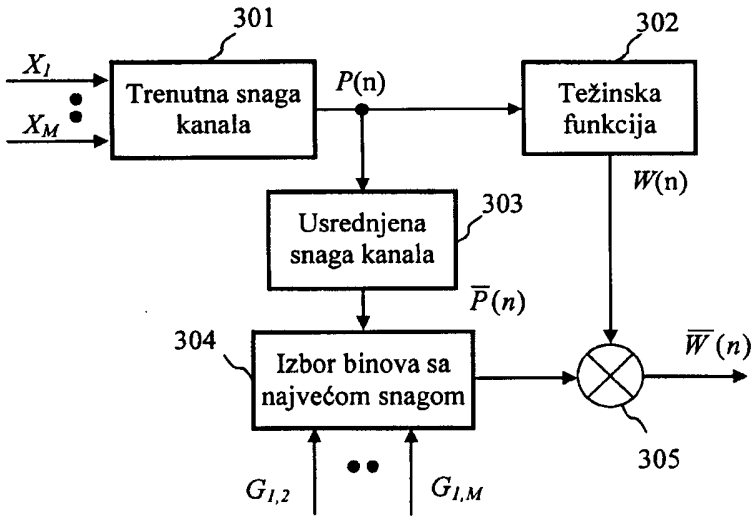
29. Postupak prema zahtevima 13 i 28 **karakterisan time,** što se ugao azimuta θ određuje na bazi aktivnosti aktuelnog govornika, pa se na izlaz sistema prosleđuje trenutna estimacija dolaznog pravca $\hat{\theta}$ u slučaju aktivnosti aktuelnog govornika, u suprotnom kada nije aktivan prosleđuje se poslednja validna estimacija azimuta $\hat{\theta}$.



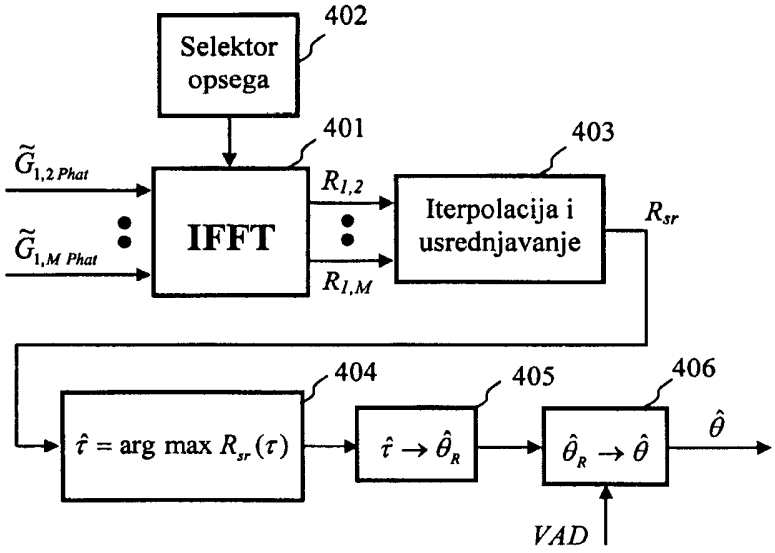
Slika 1.



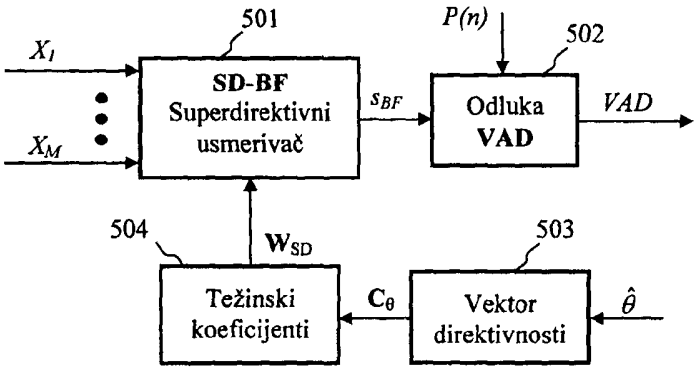
Slika 2.



Slika 3.



Slika 4.



Slika 5.

Izdaje i štampa: Zavod za intelektualnu svojinu, Beograd, Kneginje Ljubice 5