



ZAVOD ZA  
INTELEKTUALNU SVOJINU  
BEOGRAD

(51) Int. Cl.<sup>7</sup> G 05 B 19/00

(21) Broj prijave: **P- 2006/0612**  
(22) Datum podnošenja prijave: **03.11.2006.**  
(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:  
**Kukolj dr D.;**  
**Kovačević dr V.;**  
**Teslić dr N.;**  
**Pap dr I.**

(74) Zastupnik:

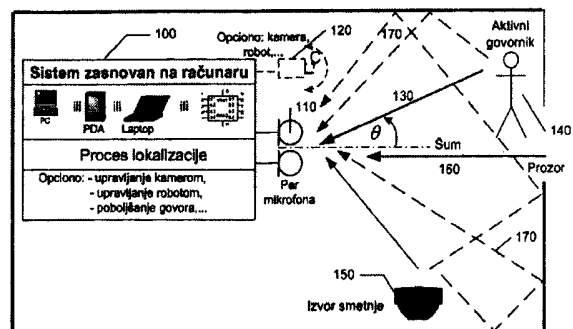
(54) Naziv: **POSTUPAK ODREĐIVANJA PRAVCA IZVORA ZVUKA ZA DVO-MIKROFONSKI SISTEM**

(51) Int. Cl.<sup>7</sup> G 05 B 19/00

(57) Apstrakt:

Pronalazak koristi dvo-mikrofonski sistem zasnovan na računaru i sadrži postupak kojim se vrši estimacija dolaznog pravca zvuka od aktivnog govornika u akustičkom ambijentu sa reverberacijom i šumom. Izloženi postupak podrazumeva da je mikrofonski par smešten u horizontalnoj ravni na definisanom rastojanju i da postupak detektuje postojanje aktivnog govornika i određuje njegov pravac, odnosno azimut ugao u odnosu na simetralu između mikrofona u horizontalnoj ravni. Pronalazak odlikuju specifični postupci: (1) predobrada segmenata jednog mikrofonskog signala sa zadatkom da se detektuju segmenti sa zvučnim glasovima jer su pogodni za korišćenje pri kros-korelaciji signala. Koristi se analiza vrednosti više obeležja u frekvencijskom domenu, tj. snaga signala u određenom spektralnom opsegu, ujednačenost spektra i spektralni centroid; (2) interpolacija podintervala vremenske kros-korelacione funkcije u cilju povećanja tačnosti dolaznog pravca; i (3) specifično inkrementalno klasterisanje vektora estimacija dolaznog pravca zvuka koje ima zadatak da poveća pouzdanost i smanji fluktuacije konačne estimacije dolaznog pravca. Sistem zasnovan na računaru u kome su implementirani postupci iz pronalaska može naći primenu u oblasti lokalizacije govornika,

odnosno upravljati robotom, poboljšati kvalitet govornog signala, upravljati kamerom i slično.



Slika 1.

## OBLAST TEHNIKE NA KOJU SE PRONALAZAK ODNOSI

Pronalazak pripada oblasti obrade akustičkog signala, ili preciznije, metodama kojima se vrši detekcija aktivnog govornika i njegovo lociranje u akustičkom ambijentu sa prisutnim šumom i reverberacijom.

## TEHNIČKI PROBLEM

Određivanje pravca izvora zvuka ili njegova lokalizacija (engl. *Direction Of Arrival - DOA*) se koristi u sistemima za komunikaciju čovek-računar putem glasa, personalnim robotima, video-telefonskim sistemima, tele-konferencijskim sistemima, itd. Određivanje pravca aktuelnog govornika u odnosu na položaj mikrofona u horizontalnoj ravni je veoma važno kod takvih sistema, pošto se ta informacija može iskoristiti za poboljšanje kvaliteta govornog signala, usmeravanje video kamere ili robota.

Postupci određivanja pravca su veoma osetljivi na smetnje prisutne u akustičkom ambijentu. Smetnje mogu nastati usled prisustva izvora nestacionarnog ili stacionarnog šuma, prisustva više govornika (izvora zvuka) koji istovremeno govore (*cocktail-party* efekat), reverberacije u prostoriji. Dodatni problem je promenljiv položaj odabranog govornika tokom vremena koji nastaje usled kretanja jednog govornika ili usled dijaloga između više govornika. Promene u akustičkom okruženju utiču na kvalitet signala dobijenih na mikrofona, tj. na karakteristike mikrofonskih signala i njihovu sličnost. Ako se zna da sistem za određivanje položaja potencijalnog izvora korisnog govornog signala mora funkcionisati u realnom vremenu, u uslovima nepoznatog akustičkog ambijenta i izvora smetnji, jasno je da se radi o složenom tehničkom problemu.

## STANJE TEHNIKE

Procena dolaznih pravaca, uz korišćenje prostorno razdvojenih mikrofona je povezana sa raznim primenama kao što su: video telefonski sistemi, telekonferencijski sistemi, "hands-free" govorna komunikacija, sistemi za poboljšanje kvaliteta i razumljivosti govornog signala, robotika, prepoznavanje govora, sistemi za praćenje i nadzor. U svim tim primenama, procena dolaznih pravaca primenjuje se u cilju lokalizacije aktivnog i dominantnog govornika (R. L. B. Jeannmes, P. Scalart, G. Faucon, and C. Beaugeant, Combined noise and echo reduction in hands-free systems: A survey, *IEEE Trans. Speech Audio Proc.*, Vol.9, pp.808–820, 2001).

Postoje više pristupa u proceni dolaznih pravaca: (a) Maksimizacija izlazne snage izlaza iz superdirektivnog mikrofonskog niza po dolaznim pravcima, (b) Visokorezolutivna procena spektra, (c) Procena vremenskih kašnjenja. Prvi pristup se retko koristi za procenu dolaznih pravaca zbog velike računске složenosti postupka. Drugi pristup, tj. visoko rezolutivna procena spektra koristi prostorno-spektralnu korelacionu matricu koja se formira na osnovu mikrofonskih signala. Veliki broj sub-optimalnih tehnika sa redukovanom kompleksnošću iz ove klase je dobro poznat. Tu spadaju: metod minimalne varijanse, metod minimalne norme, metod višestruke klasifikacije signala (MUSIC), itd.

Metode procene vremenskih kašnjenja (*Time-delay Estimation, TDE*) bazirane su na proceni vremenskih kašnjenja između mikrofonskih signala i u praksi se najviše koriste (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.). Bazira se na lokalizaciji maksimuma kros-korelacione funkcije između para mikrofonskih signala. Metoda generalizovane kros-korelacije je poznata metoda iz klase TDE metoda za procenu dolaznih pravaca i zasniva se na lokalizaciji maksimuma kros-korelacione funkcije između ponderisanih mikrofonskih signala. Funkcija se ponderiše zbog povećanja robusnosti algoritma procene dolaznih pravaca na prisustvo šuma i odjeka u prostori. Dve najčešće korišćene težinske funkcije su: funkcija raspodele maksimalne verodostojnosti (*ML, Maximum Likelihood*) i funkcija fazne transformacije (*PHAT, Phase Transform*). Dok funkcija maksimalne verodostojnosti ističe signal na frekvencijama gde je odnos signala i šuma veliki (*SNR, Signal to Noise Ratio*), primenom funkcije fazne transformacije, poravnava se amplitudno-spektralna karakteristika signala, pri čemu se takođe sistem čini invarijantnim na snagu mikrofonskih signala, jer se kros-korelacija u spektru normira.

Najopštiji i često korišćeni pristup u lociranju izvora zvuka koristi estimaciju vremenske razlike dolaska zvučnog talasa na dva ili više mikrofona. Vremenska razlika kod

dva signala predstavlja malo kašnjenje u vremenu što odgovara faznoj razlici u frekventnom domenu. Određivanje vremena kašnjenja se postiže nalaženjem maksimuma kros-korelacije ovih signala. Pristup vremenskog kašnjenja se zasniva na nekoliko pretpostavki: da akustički talas dolazi od izvora zvuka do mikrofona po jednoj, direktnoj putanji; da nema drugih izvora zvuka ili da su znatno slabiji; da je odziv mikrofona približno isti. Poslednja pretpostavka se zadovoljava korišćenjem identičnih mikrofona, iste orijentacije i malog međusobnog rastojanja u odnosu na rastojanje izvora zvuka. Međutim, pojava reverberacije u prostoriji, prisustvo više izvora zvuka i šuma znatno smanjuju tačnost ovog pristupa. U tu svrhu je razvijeno više različitih pristupa za povećanje robusnosti lokalizacije.

Postoji veći broj različitih rešenja koja su patentirana i koja rešavaju problem korišćenjem samo dva mikrofona ili korišćenjem mikrofonskog niza ili u kombinaciji sa digitalnom kamerom. Na primer, EP objavljena patentna prijava 1 286 175 A2, prijavljena 1. avgusta 2002., sa naslovom „Robust talker localization in reverberant environment“, daje rešenje za lokalizaciju govornika u reverberantnoj sobi pomoću usmerene karakteristike mikrofonskog niza, zatim EP objavljena patentna prijava 1 205 762 prijavljena 7. juna 2000., sa naslovom „Method and apparatus for determining sound source“, daje rešenje za lokalizaciju izvora zvuka na osnovu zvuka dobijenog sa dva mikrofona i slike sa kamere, pri čemu se lokalizacija zvukom zasniva na razlici između faza i razlici između intenziteta dvo-kanalnog signala, zatim US objavljena patentna prijava 6 999 593 prijavljena 28. maja 2003., sa naslovom “System and process for robust sound source localization”, daje rešenje za mikrofonski niz kombinovanjem težinske kros-korelacije i podešene usmerene karakteristike parova mikrofonskog niza, zatim US objavljena patentna prijava 6 970 796 prijavljena 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 sa statističkim klasterisanjem estimacije lokacija, zatim US objavljena patentna prijava 6 826 284 prijavljena 4. februara 2000., sa naslovom “Method and apparatus for passive acoustic source localization for video camera”, daje rešenje koje na osnovu dva mikrofona locira izvor zvuka koristeći kombinaciju algoritma adaptivne dekompozicije sopstvenih vrednosti i metode najmanjih kvadrata, zatim US objavljena patentna prijava 6 005 610 prijavljena 23. januara 1998., sa naslovom “Audio-visual object localization and tracking system and method therefore”, daje integralno audio-vizuelno rešenje za lokalizaciju sa primarnim težištem na vizuelna obeležja objekta, kao i US objavljena patentna prijava 6 952 672 prijavljena 25. aprila 2001., sa naslovom “Audio source

position detection and audio adjustment”, daje rešenje koje kombinuje korišćenje mikrofona sa infracrvenim senzorima.

## **IZLAGANJE SUŠTINE PRONALASKA**

Predmet ovog pronalaska je postupak kojim se određuje pravac aktivnog govornika u odnosu na poziciju mikrofona. Postupak podrazumeva da je par mikrofona smešten u horizontalnoj ravni na definisanom rastojanju i da postupak detektuje postojanje aktivnog govornika i određuje njegov pravac, odnosno azimut u odnosu na simetralu između mikrofona u horizontalnoj ravni. Pri tome u akustičkom ambijentu mogu biti prisutne brojne smetnje. One nastaju kao posledica direktnog zvučnog talasa od jednog ili više izvora šuma ili izvora drugih smetnji koji se nalaze u prostoriji (muzika, kućni uređaji i sl.) ili van nje (saobraćajna buka, na primer), kao i svi reflektovani talasi koji potiču od svih izvora zvukova, uključujući i govornika, a koji nastaju usled reverberacije prostorije. Treba naglasiti da su izvori zvukova u prostoriji najčešće nestacionarni.

Sušтина pronalaska je u specifičnoj obradi mikrofonskog dvo-kanalnog signala dobijenog u akustičkom ambijentu prostorije. Prva specifičnost pronalaska je u procesu obrade segmenata jednog mikrofonskog signala sa zadatkom da se detektuju segmenti sa zvučnim glasovima jer su pogodni za korišćenje u postupku korelisanja signala. Takve segmente odlikuje veća periodičnost – pravilnost od zašumljenih glasova i šumova. Detekcija zvučnih glasova se odvija na signalu filtriranim pomoću dva pojasno-propusna filtra sa različitim parametrima. U procesu detekcije zvučnih segmenata signala se analiziraju vrednosti više obeležja u frekvencijskom domenu i to: snaga signala u određenom spektralnom opsegu, ujednačenost spektra i spektralni centroid.

Nad segmentima dvo-mikrofonskog signala koji su detektovani kao zvučni se primenjuje generalizovana kros-korelacija sa težinskom funkcijom fazne transformacije – PHAT funkcijom, pomoću koje se estimira mogući dolazni pravac direktnog zvučnog talasa od govornika. Međutim, zbog prisustva visokog nivoa reverberacije ili šuma, ovaj postupak nije pouzdan. Zato je robusnost procene dolaznog pravca signala poboljšana uvođenjem obrade koja se sastoji iz dva specifična procesa: interpoliranja podopsega dolaznog pravca od interesa u višestruko više tačaka; i inkrementalnog klasterisanja dobijenih estimacija dolaznih pravaca organizovanih u vremensku seriju sa pomičnim vremenskim prozorom.

Povećanje tačnosti estimacije dolaznog pravca zvuka pomoću interpolacije podintervala vremenske kros-korelacione funkcije je sledeća specifičnost pronalaska. Interpolacija se vrši u okolini inicijalno dobijenog maksimuma sa fazno transformisanom generalizovanom kros-korelacijom. Nakon određivanja podintervala i *spline* interpolacije, vreme kašnjenja signala u podrezoluciji se nalazi pomoću specifičnih kriterijuma.

Poslednja specifičnost je u inkrementalnom klasterisanju estimacija dolaznog pravca zvuka. Postupak je uveden u cilju povećanja pouzdanosti i smanjenja fluktuacija konačne estimacije dolaznog pravca. Generalno, klaster analiza podrazumeva grupisanje uzoraka (*pattern clustering*) u određeni broj homogenih grupa (*clusters*) na bazi odabrane mere sličnosti između uzoraka. Uzorci grupisani u isti klaster treba da budu slični jedni drugima, dok to ne treba da važi za uzorke iz različitih klastera. U slučaju procene dolaznog pravca zvuka, klaster analiza se obavlja nad vremenskom serijom u kojoj svaka vrednost predstavlja vremensko kašnjenje koje je rezultiralo ranijom obradom nad funkcijom korelacije i njenom interpolacijom. Klaster analiza serije estimacija dolaznog pravca se obavlja specifičnim algoritmom inkrementalnog klasterisanja.

#### KRATAK OPIS SLIKA I NACRTA

**Slika 1** – prikazuje akustički ambijent u kojem se nalazi dvo-mikrofonski sistem za određivanje dolaznog pravca zvuka u cilju lokalizacije govornika.

**Slika 2** – prikazuje fokusirano deo sistema za određivanje dolaznog pravca zvuka, koji obuhvata pred-obradu sa ciljem da odredi zvučne segmente signala. Slika prikazuje i centralni deo postupka koji sadrži kros-korelaciju PHAT metodom.

**Slika 3** – prikazuje deo sistema za određivanje dolaznog pravca zvuka, koji se odvija nakon procene pravca PHAT metodom i obuhvata post-obradu. Post-obrada sadrži inkrementalno klasterisanje estimacije dolaznog pravca zvuka u podrezoluciji korelacione funkcije.

#### DETALJAN OPIS PRONALASKA

Ovaj pronalazak opisuje dvo-mikrofonski sistem i postupak kojim se poboljšava estimacija dolaznog pravca zvuka od aktivnog govornika u akustičkom ambijentu sa reverberacijom i šumom.

**Slika 1** prikazuje sistem zasnovan na računaru u odgovarajućem akustičkom ambijentu. Namena sistema **100** je da tačnije i pouzdanije odredi dolazni pravac zvuka od aktivnog govornika u odnosu na par identičnih i istovetno orijentisanih mikrofona **110**. Tačnije, određuje se azimut ugao  $\theta$  između pravca dolaska direktnog zvučnog talasa **130** i simetrale između mikrofona u horizontalnoj ravni.

Postupci iz ovog pronalaska se mogu softverski implementirati na različitim tipovima procesora (**100**): PC, PDA, DSP itd. Takođe, moguće ih je implementirati korišćenjem specifičnih integrisanih kola (ASIC), programabilnih logičkih kola (PLD ili FPGA) i slično. Postupci koji obuhvataju ovaj pronalazak omogućavaju sistemu zasnovanom na računaru **100** da zavisno od primene: upravlja robotom, poboljšava kvalitet govornog signala, upravlja kamerom **120** i slično. Postupci iz ovog pronalaska su pod kontrolom većeg broja objašnjenih parametara koji omogućavaju podešavanje i optimizaciju rešenja za različite primene.

Sistem **100** funkcioniše u akustičkom ambijentu u kojem pored govornika **140** mogu biti prisutne brojne smetnje prouzrokovane uticajem drugih internih izvora zvuka **150**, spoljašnjih izvora zvuka (buka saobraćaja, na pr.), difuznog šuma **160**, kao i usled reflektovanih talasa od svih izvora zvuka **170**.

Na **Slici 2** se vidi da se signali  $x_L$  i  $x_R$  sa mikrofonskog para **110** dele u bloku **200** na segmente od  $N$  odbiraka koji su 50% preklapljeni u vremenu, a potom se svaki segment množi sa Hamming-ovom prozorskom funkcijom. Pri tome se na oba kraja segmenta dodaje po  $N/2$  nula radi slabljenja efekata transformacije na krajevima segmenta i povećanja njene rezolucije. Na kraju postupka u bloku **200** se na segmentima signala primeni diskretna Furijeova transformacija na konačnom vremenskom intervalu (*short-time discrete Fourier transform*). Izlazni segmenti signala  $X_L$  i  $X_R$  se koriste u daljem procesu. Treba naglasiti da postupak pred-obrađe gde se obavlja detektovanje zvučnih glasova u segmentima signala, koristi samo jedan, bilo koji signal. Ovde je odabran signal  $X_L$ .

VAD blok **320** je detektor govorne aktivnosti baziran na jednostavnom algoritmu određivanja energije kratkotrajnih vremenskih segmenata, prethodno filtriranih pojasnim filtrom (*band-pass*) u bloku **310** definisanim sa granicama  $LowL=300\text{Hz}$  i  $HighL=1500\text{Hz}$ . Pošto je energija  $E$  vremenskog segmenta diskretnog signala jednaka energiji njegovog spektra, adekvatan izraz za energiju filtriranog signala korišćen u VAD bloku **320** je:

$$E = \sum_{i=LowL}^{HighL} X(i)^2, \quad (1)$$

gde je  $X(i)$  komponenta  $i$ -te frekvencije u frekventnom opsegu od  $(LowL \div HighL)$ .

Zvučni glasovi imaju obično višu energiju u razmatranom frekventnom opsegu od

bezvučnih ili šuma, zbog čega se u bloku 320 proverava da li za dati vremenski segment važi uslov  $E > T_{NVAD}$ , gde je  $T_{NVAD}$  prag energije koji se sporo adaptira u datim granicama sa rekurzivnim filtrom prvog reda, tj:  $T_{NVAD} = 0.99T_{NVAD} + 0.01E$ ,  $T_{NVAD}^{LowL} \leq T_{NVAD} \leq T_{NVAD}^{HighL}$ , sa parametrima definisanim u bloku 370. VAD algoritamski blok 320 detektuje postojanje zvučne govorne aktivnosti. Ako ovaj uslov nije ispunjen, tekući segment se dalje ne razmatra i čeka se sledeći segment signala.

Ako je prethodni uslov ispunjen, ulazni segment signala se filtrira pojasnim filtrom 330 u frekvencijskom opsegu od 300Hz do 4kHz. Za svaki filtrirani segment se računaju dva spektralna obeležja: ujednačenost spektra ( $SF$ , *spectral flatness*) i spektralni centroid ( $SC$ , *spectral centroid*). Pomoću ova dva obeležja se detektuju zvučni glasovi.

Ocena nivoa ujednačenosti spektra  $SF$  predstavlja meru zašumljenosti, dekorelisanosti i ujednačenosti spektra ili jednog njegovog dela. Računa se u bloku 340 kao odnos između geometrijske i aritmetičke sredine energije spektra signala, odnosno,

$$SF = \frac{\left( \prod_{i=1}^{N/2} X(i) \right)^{1/N/2}}{\frac{1}{N/2} \sum_{i=1}^{N/2} X(i)} \quad (2)$$

gde je  $X(i)$  amplituda  $i$ -te frekvencije u frekventnom opsegu do  $N/2$ , a  $N/2$  odgovara frekvenciji od 4kHz.

U bloku 350 se računa spektralni centroid  $SC$ . Spektralni centroid  $SC$  predstavlja težište spektra tekućeg segmenta i računa se kao:

$$SC = \frac{\sum_{i=1}^{N/2} f_i |X(i)|}{\sum_{i=1}^{N/2} |X(i)|} \quad (3)$$

gde je  $f_i$   $i$ -ta frekvencija odgovarajuće spektralne amplitude  $X(i)$ .

Izračunate vrednosti obeležja  $SF$  i  $SC$  se proveravaju u bloku za odlučivanje 360. Za meru ujednačenosti spektra su predefinisane dve granične vrednosti: viša –  $SF^h$  i niža –  $SF_l$ . Ako je izračunata vrednost  $SF$  mere manja od  $SF_l$ , onda je tekući segment zvučan, a ako je  $SF$  mera veća od  $SF^h$ , onda tekući segment nije zvučni glas. Ako nijedan od dva uslova nije ispunjen, onda je zvučnost tekućeg segmenta jednaka prethodnom segmentu. Ocena spektralnog centroida obično daje niže vrednosti za zvučne glasove. Zbog toga se proverava da li je vrednosti spektralne centroide datog segmenta ispod granične vrednosti praga, tj.  $SC < T_{hSC}$ . Ako ovaj uslov nije ispunjen, segment ne sadrži zvučni glas. Samo u slučaju ispunjenja oba uslova donosi se konačna odluka da tekući segment odgovara zvučnom glasu i



promenljivom  $\text{Ind}_{\text{voiced}}$  se aktivira sledeći blok obrade 400. Svi parametri za detekciju zvučnih segmenata su definisani u bloku 370 (na pr.:  $T_{hVAD} \approx 0.3$ ,  $SF_l = 0.45$ ,  $SF^h = 0.49$ ,  $T_{hSC} = 0.2$ ).

Centralno mesto postupka zauzima blok 400 za proračun kros-korelacije segmenata mikrofonskih signala  $x_L$  i  $x_R$ , tj. njihovih Furijeovih transformacija  $X_L$  i  $X_R$ . Korišćena je varijanta generalizovane kros-korelacije (GCC) sa funkcijom fazne transformacije (PHAT). PHAT verzija GCC metode procenjuje vremensko kašnjenje signala po sledećoj relaciji:

$$\hat{\tau}_{\text{PhatCC}} = \arg \max_n \hat{\psi}_{\text{PhatCC}}[n] \quad (4)$$

gde je:

$$\hat{\psi}_{\text{PhatCC}}[n] = \sum_{k=0}^{N-1} \phi_{\text{PHAT}}[k] S_{x_L, x_R}[k] e^{j \frac{2\pi nk}{N}}, \quad (5)$$

pri čemu je  $N$  dužina segmenta,  $S_{x_L, x_R}[k] = X_L[k] X_R^*[k]$  je kros-spektar mikrofonskih signala (za determinističke signale), oznake  $X_l[k], l \in \{L, R\}$  predstavljaju diskretnu Furijeovu transformaciju signala levog i desnog mikrofona, dok  $\phi_{\text{PHAT}}[k]$  predstavlja spektralnu težinsku funkciju fazne transformacije PHAT. Iz priloženih relacija se vidi da  $\hat{\psi}_{\text{PhatCC}}[n]$  predstavlja procenu kros-korelacione funkcije između mikrofonskih signala, tako da se traženjem argumenta te funkcije koji je maksimizira, dobija procena kašnjenja između signala.

PHAT težinska funkcija je definisana kao:

$$\phi_{\text{PHAT}}[k] = \frac{1}{|S_{x_L, x_R}[k]|}. \quad (6)$$

Vidi se da se primenom PHAT težinske funkcije normalizuje spektar po modulu, pri čemu onda u filtriranom kros-spektaru figurišu samo fazna kašnjenja mikrofonskih signala, čime je procena invarijantna na snagu signala. Na taj način, invarijantnost na snagu signala (uz pretpostavku da je šum jednak nuli) omogućava PHAT funkciji osobinu da zavisi samo od fizičke postavke sistema, tj. položaja izvora zvuka i mikrofonskog para i od akustičnog okruženja.

Nakon inverzne Furijeove transformacije izraza (5), realni deo determiniše talasni oblik funkcije kros-korelacije u opsegu od  $-L$  do  $+L$  odbiraka. Broj  $L$ , kao najveće kašnjenje u odbircima između dva kanala, određen je rastojanjem između mikrofona  $d_{\text{mic}}$  i periodom odabiranja  $fs$ , prema relaciji:  $L = fs d_{\text{mic}} / (2c)$ , gde je sa  $c$  označena brzina zvuka ( $c = 342 \text{ m/s}$ ). Konačni izlaz bloka 400 je *phat* funkcija kros-korelacije i vremensko kašnjenje  $\tau_{\text{cc}}$  koje odgovara maksimumu *phat* funkcije  $\max_{\tau_{\text{cc}}}$ . Vrednosti *phat*,  $\tau_{\text{cc}}$  i  $\max_{\tau_{\text{cc}}}$  se prosleđuju daljoj

post-obradi u bloku 500.

Prisustvo šuma i drugih promena u akustičnom okruženju smanjuju pouzdanost i tačnost estimacije dolaznog pravca od izvora zvuka. Da bi se povećala tačnost estimacije, vrši se interpolacija *phat* funkcije, čime se postiže povećanje rezolucije procene azimut ugla. Realizovanom interpolacijom se povećava rezolucija estimacije 4 do 8 puta, opciono. Dobijena *phat* funkcija se razmatra u intervalu  $(-L, L)$  definisanim sa gore navedenim parametrom  $L$ , koji sadrži  $2L+1$  tačaka.

Slika 3 detaljno prikazuje tok obrade naznačen blokom 500 u Slici 2. Na osnovu određenog maksimuma amplitude *phat* funkcije  $max_{TCC}$  iz bloka 400, u okolini tog maksimuma se u bloku 510 *spline* interpolacijom povećava broj tačaka. Pri tome se u bloku 510 javljaju tri slučaja:

(1) Locirani maksimum funkcije je dovoljno udaljen od krajnje leve ivice  $(-L)$  i desne ivice  $(+L)$  intervala: Pojam 'dovoljno udaljen' je definisan parametrom  $w$  koji određuje granice podintervala nad kojim se vrši interpolacija. To znači da je maksimalna dužina interpolacionog intervala  $2w+1$  ( $-w$  tačaka ulevo od lokacije maksimuma,  $w$  tačaka udesno i jedna u maksimumu).

(2) Locirani maksimum *phat* funkcije je blizu leve granice intervala: Tada se uzimaju tačke od između  $-L$  i lociranog maksimuma i  $w$  tačaka sa desne strane maksimuma.

(3) Locirani maksimum *phat* funkcije je blizu desnoj granici intervala: Razmatra se  $w$  tačaka levo od maksimuma i tačke od lociranog maksimuma do desne ivice intervala.

Rezultat bloka 510 je interpolirani podinterval *phat* funkcije  $V$ .

Sledeći blok 520 nalazi maksimalnu vrednost  $max_{PHAT}$  u interpoliranom podintervalu  $V$ .

Za razliku od drugih pristupa gde se estimacija dolaznog pravca određuje prostom lokacijom maksimuma; ponderisanjem prvog i drugog maksimuma; odnosa maksimuma i srednje vrednosti i sl.; ovde se u bloku 530 estimacija dolaznog pravca vrši sa dva različita kriterijuma, od kojih bar jedan mora biti ispunjen:

1. kriterijum: Da li je maksimum interpoliranog podintervala *phat* funkcije  $max_{PHAT}$  veći od adaptivnog praga  $T_{hPEAKadapt}$ . Rekurzivna adaptacija ovog praga se obavlja sa filtrom (*moving average*) oblika:

$$T_{hPEAKadapt} = 0.7 T_{hPEAKadapt} + (1-0.7) max_{PHAT} \quad (7)$$

2. kriterijum: Proverava se da li maksimum interpoliranog podintervala *phat* funkcije  $max_{PHAT}$  ispunjava uslov:

$$\max_{PHAT} > T_{hPEAK2} \wedge ((\max_{PHAT} / V(1)) > 4 \wedge (\max_{PHAT} / V(n)) > 4) \quad (8)$$

gde je  $T_{hPEAK2}$  konstantni prag ( $T_{hPEAK2}=0.06$ ), a  $V(1)$  i  $V(n)$  su vrednosti prve i poslednje tačke interpoliranog podintervala.

U slučaju da je jedan od ova dva uslova ispunjen, nova vrednost finalne estimacije dolaznog pravca  $\tau_{PHAT}$  odgovara tekućem maksimumu  $\max_{PHAT}$  interpoliranog podintervala  $phat$  funkcije.

Estimacija dolaznog pravca izvora zvuka iz bloka **530**  $\tau_{PHAT}$  se smešta u bafer **D** (blok **540**), koji predstavlja pomični prozor u vremenskoj seriji sukcesivnih estimacija. Elementi bafera **D** iz bloka **540** se podvrgavaju postupku klaster analize sa ciljem da se iz potencijalno uočenih klastera izdvoje pokazatelji koji će uticati na povećanje tačnosti i pouzdanosti estimiranih dolaznih pravaca zvuka.

Algoritam klaster analize se odvija u bloku **550**. Startuje kada ulazni bafer **D** iz bloka **540** sadrži  $K$  poslednjih validnih procena dolaznog pravca. Ako je bafer već pun, nova procena iz bloka **530** potiskuje najstariju. Broj postojećih klastera je inicijalno postavljen na jedan ( $G = 1$ ) i centar tog klastera je postavljen na srednju vrednost ulaznog bafera **D**.

Tok klaster analize se kontoliše sa nekoliko predefinisanih parametara smeštenih u blok **560**. Svaki klaster je određen svojim centrom (verovatni dolazni pravac zvuka), rasipanjem i vremenskim trajanjem. Trajanje klastera je određeno učestanošću pojavljivanja vrednosti estimacija dolaznog pravca koje mu pripadaju, tj. vremenom kada se poslednje pridruživanje estimacije dolazanog pravca nekom klasteru desilo.

Formiranje novog klastera koji odgovara novom dolaznom pravcu se kontroliše pomoću dva prethodno definisana praga  $Th^{C1}$  i  $Th^{C2}$  iz bloka **560**. Pomoću praga  $Th^{C1}$  se identifikuju fluktuacije vrednosti estimacije dolaznog pravca koje se nalaze u pomičnom baferu **D** bloka **540**. One odgovaraju relativno sporij promeni dolaznog pravca koja može odgovarati efektu govornika u pokretu. Pri tome prag  $Th^{C1}$  predstavlja širinu nepreklapajućih klastera, ne dozvoljavajući male fluktuacije finalne estimacije oko vrednosti stvarnog dolaznog pravca, a unutar granica definisanih sa  $Th^{C1}$ . Prag  $Th^{C2}$  je vezan za tekuću estimaciju dolaznog pravca i može odgovarati nagloj promeni, usled prelaska estimacije pravca sa jednog na drugog aktivnog govornika.

Pored pragova  $Th^{C1}$  i  $Th^{C2}$ , blok **560** sadrži definisani maksimalni broj klastera  $G_{max}$  i maksimalnu vrednost trajanja klastera  $T_{max}$ . Ovi parametri određuju vreme trajanja nekog klastera, tj. odgovarajućeg dolaznog pravca zvuka. Trajanje klastera je određeno vremenom proteklim od njegovog poslednjeg korišćenja. U slučaju kada se formira novi klaster, uvek se

proverava broj postojećih klastera i trajanje 'najstarijeg' klastera. Dobre vrednosti parametara iz bloka 560 su:  $Th^{C1}=2$ ,  $Th^{C2}=5$ ,  $G_{max}=3$ ,  $T_{max}=100$ .

Za svaku procenu dolaznog pravca zvuka iz bloka 540, u bloku 550 klaster analize se izvrše sledeći koraci:

1. Računa se srednja vrednost  $d$  i varijansa  $s$  poslednjih  $K$  vrednosti ulaznog bafera  $D$  kao:

$$d = \frac{1}{K} \sum_{i=1}^K D(i), \quad s = \frac{1}{K} \sum_{i=1}^K (D(i) - d)^2. \quad (9)$$

2. Ako je vrednost varijanse  $s$  ispod prethodno definisanog praga  $Th^{C1}$ , trenutna srednja vrednost  $d$  elemenata bafera  $D$  se dodeljuje najbližem postojećem  $w$ -tom klasteru  $m(w)$ , tj.  $d \in m(w)$ , ako važi  $\|d - m(w)\| < \|d - m(g)\|$ ,  $g=1, \dots, G$ , i  $\|d - m(w)\| < T^{C1}$ , gde je  $\|\cdot\|$  Euklidska norma i gde je prag  $Th^{C1}$  iz bloka 560 izabran tako da nema preklapanja između klastera. Ako se sa  $p$  predstavi vektor sa trajanjima svih klastera, onda komponenta vektora  $p$  za dato  $w$  (gde je  $w$  indeks klastera koji je primio  $d$ ) postaje  $p(w) = k$ .

3. Ukoliko je ipak vrednost varijanse  $s$  iznad praga  $Th^{C1}$ , ili nije ispunjen uslov najbližeg klastera  $\|d - m(w)\| < T^{C2}$ , tada se formira novi klaster sa centrom  $m(G) = d$  i izvrši se operacija  $G = G+1$ ,  $G \leq G_{max}$ , gde je  $G_{max}$  maksimalno dozvoljen broj klastera. Takođe, ako je ispunjen uslov  $(G > G_{max}) \vee (\exists g, (k - p(g)) > T_{max})$ , tada se "najstariji" klaster uklanja iz skupa klastera. Parametri  $G_{max}$  i  $T_{max}$  su iz bloka 560.

4. Izračunavaju se nove varijanse elemenata bafera  $D$  u odnosu na centar svakog klastera  $m(g)$ ,  $g=1, \dots, G$ . Vrednost centra klastera za koji je ova vrednost najmanja predstavlja izlaz bloka 550 u obliku konačne estimacije dolaznog pravca  $\tau_{fin}$ .

Izlaz  $\tau_{fin}$  iz bloka 550 klaster analize je u formi vremenskog kašnjenja između mikrofonskih signala  $x_L$  i  $x_R$ . U bloku 570 se estimacija  $\tau_{fin}$  transformiše u azimut ugao  $\theta$ , koji je pored kašnjenja  $\tau_{fin}$  u funkciji od brzine zvuka, rastojanja mikrofona i frekvencije odabiranja.

Prethodni opis je detaljno predstavio ceo postupak detekcije i lokalizacije govornika, ili preciznije, određivanja dolaznog pravca govornog signala. Postupak odlikuju faze pred-obrađe i post-obrađe koje su povećale tačnost i pouzdanost postupka. Ceo postupak se može koristiti kao samostalan sistem za praćenje govornika, ili u okviru nekog još složenijeg sistema kao njegov sastavni deo, na primer kod: sistema za navigaciju robota, upravljanje kamerom kod *hands-free* sistema ili sistema za poboljšanje govornog signala. Takođe, treba napomenuti da postupci iz: (1) pred-obrađe koja sadrži analizu više obeležja u frekventijskom

domenu; i (2) post-obrade koja obuhvata inkrementalno klasterisanje estimacija dolaznog pravca interpoliranih u sub-rezoluciji, mogu pojedinačno naći i druge forme primene. Detaljni opis ovog pronalaska omogućavaju stručnjaku iz oblasti obrade signala da generičke principe ovog pronalaska uspešno implementira u okviru odabrane primene pri čemu se ne izlazi iz okvira ovog pronalaska.

**PATENTNI ZAHTEVI**

1. Postupak određivanja pravca izvora zvuka za dvo-mikrofonski sistem **karakterisan time** što kao izvor zvuka koristi signale sa dva identična mikrofona iste orijentacije i malog rastojanja gde se vrši korelacija ulaznih signala kojoj prethodi specifična pred-obrada signala, a potom post-obradom dodatno poboljšava estimaciju dolaznog pravca govornog signala u odnosu na simetralu para mikrofona.

2. Postupak prema zahtevu 1 **karakterisan time** što se pred-obrada signala odvija na jednom odabranom mikrofonskom signalu, odnosno na segmentima odabranog signala.

3. Postupak prema zahtevu 2 **karakterisan time** što detektuje segmente sa zvučnim glasovima pomoću više spektralnih obeležja, detekcija zvučnih glasova se odvija na signalu filtriranom pomoću dva pojasno propusna filtra sa različitim parametrima.

4. Postupak prema zahtevu 3 **karakterisan time** što analizira više obeležja u frekvencijskom domenu i to snagu signala u filtriranom spektralnom opsegu (320), ravnost spektra (340) i težište spektra (350).

5. Postupak prema zahtevu 1 **karakterisan time** što se pomoću post-obrade koja ima dva sukcesivna koraka i to interpoliranje podopsega pravca od interesa sa četiri ili osam puta više tačaka (510) i inkrementalno klasterisanje dobijenih estimacija organizovanih u vremensku seriju sa pomičnim vremenskim prozorom (550).

6. Postupak prema zahtevu 2 **karakterisan time** što se postupak detekcije zvučnih segmenata postupkom analize obeležja koristeći tri frekvencijska obeležja odvija na jednom (bilo kojem) mikrofonskom signalu.

7. Postupak prema zahtevu 6 **karakterisan time** što se filtriranje ulaznog signala vrši sa pojasno propusnim filtrom sa opsegom 300Hz-1500Hz (310).

8. Postupak prema zahtevu 6 **karakterisan time** što se vrši provera snage filtriranog segmenta signala u odnosu na adaptivni sporo promenljivi prag, tako da ako je snaga signala iznad vrednosti praga, postupak detekcije se nastavlja, u suprotnom detekcija se odmah prekida i čeka se novi segment.

9. Postupak prema zahtevu 6 **karakterisan time** što se filtriranje ulaznog signala vrši sa pojasno propusnim filtrom u frekvencijskom opsegu od 300 Hz do 4 KHz (330).

10. Postupak prema zahtevima 6 i 9 karakterisan time što se proračun (340) i provera (360) vrednosti ujednačenosti spektra filtriranog segmenta signala  $X_L$  u odnosu na dva predefinisana praga, višeg i nižeg, vrši tako da ako je izračunata vrednost ujednačenosti spektra manja od nižeg praga onda je tekući segment potencijalno zvučan, a ako je vrednost mera veća od višeg praga onda tekući segment nije zvučan, ako nijedan od dva uslova nije ispunjen onda je zvučnost tekućeg segmenta jednaka zvučnosti prethodnog segmenta.

11. Postupak prema zahtevu 10 karakterisan time što se provera za proračun (350) i proveru (360) vrednosti spektralne centroide za filtrirani segment signala  $X_L$  da li je ispod vrednosti praga  $T_{hSC}$ , vrši tako da ako ovaj uslov nije ispunjen, segment nije zvučan.

12. Postupak prema zahtevima 10 i 11 karakterisan time da u slučaju da su obe provere rezultirale sa potencijalno zvučnim segmentom, onda je detektovan zvučni segment signala.

13. Prvi deo post-obrade prema zahtevu 5 ima zadatak da sa interpoliranjem podintervala funkcije oko inicijalno estimiranog pravca sa četiri ili osam puta više tačaka (po izboru, opciono) poveća tačnost estimacije i karakterisan time što određuje podinterval za interpolaciju i nalazi maksimum podintervala funkcije kros-korelacije *phat* po specifičnim kriterijumima, pri čemu određivanje podintervala nad kojim će se vršiti *spline* interpolacija (510) razlikuje tri slučaja, zavisno od toga gde se nalazi inicijalno određeni maksimum funkcije *phat*, odnosno od odnosa širine interpolacionog podintervala  $(-w, +w)$  prema granicama određenosti funkcije *phat*  $(-L, +L)$ , dok se estimacija dolaznog pravca vrši pomoću dva različita kriterijuma (530): (1)-da li je maksimum interpoliranog podintervala *phat* funkcije  $max_{PHAT}$  veći od rekurzivno adaptivnog praga  $T_{hPEAKadapt}$  po obrascu:

$$T_{hPEAKadapt} = 0.7 T_{hPEAKadapt} + (1-0.7) max_{PHAT},$$

i (2)-da li maksimum interpoliranog podintervala *phat* funkcije  $max_{PHAT}$  ispunjava uslov:

$$max_{PHAT} > T_{hPEAK2} \wedge ((max_{PHAT} / V(1)) > 4 \wedge (max_{PHAT} / V(n)) > 4)$$

gde je  $T_{hPEAK2}$  konstantni prag ( $T_{hPEAK2}=0.06$ ), a  $V(1)$  i  $V(n)$  su vrednosti prve i poslednje tačke interpoliranog podintervala, tako da u slučaju da je jedan od ova dva uslova ispunjen, nova vrednost finalne estimacije dolaznog pravca  $\tau_{PHAT}$  odgovara tekućem maksimumu  $max_{PHAT}$  interpoliranog podintervala *phat* funkcije.

14. Postupak prema zahtevu 13 karakterisan time da se nastavak post-obrade sastoji iz klasterisanja estimacija dolaznog pravca zvuka (550).

15. Postupak prema zahtevima 5 i 13 **karakterisan time** što se vrednosti estimacije posmatraju kao vremenska serija u kojoj se vrednosti grupišu u klasterne koji odgovaraju dolaznom pravcu zvuka.

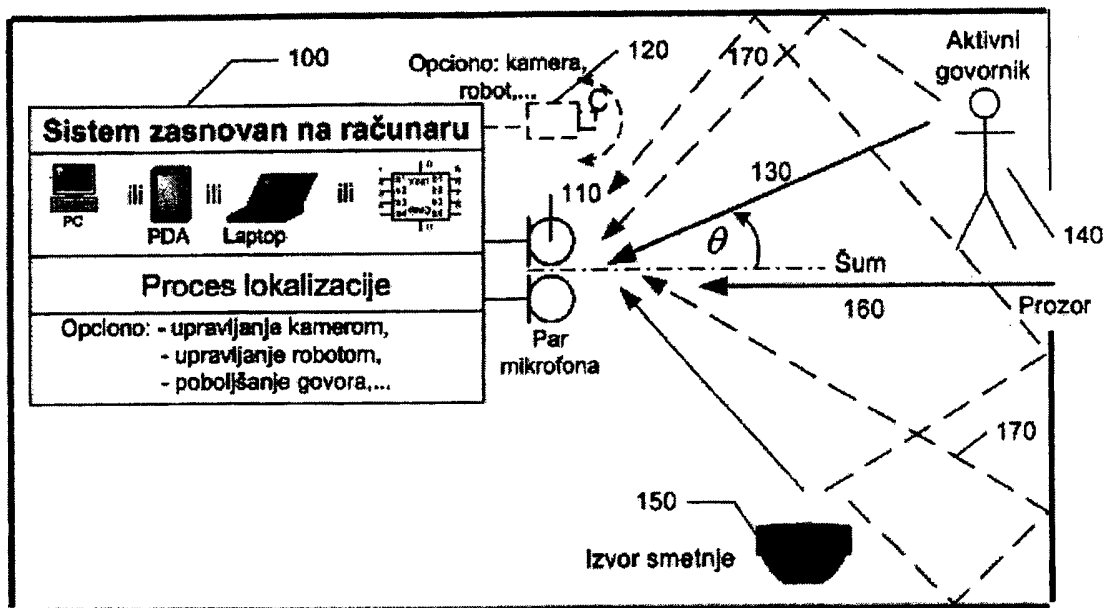
16. Postupak prema zahtevu 15 **karakterisan time** što je svaki klaster određen svojim centrom koji predstavlja verovatni dolazni pravac zvuka, rasipanjem i vremenskim trajanjem.

17. Postupak prema zahtevu 16 **karakterisan time** što je trajanje klastera određeno učestanošću ponavljanja vrednosti estimacija dolaznog pravca koje mu pripadaju kao vremenu kada se poslednje pridruživanje vrednosti nekom klasteru desilo.

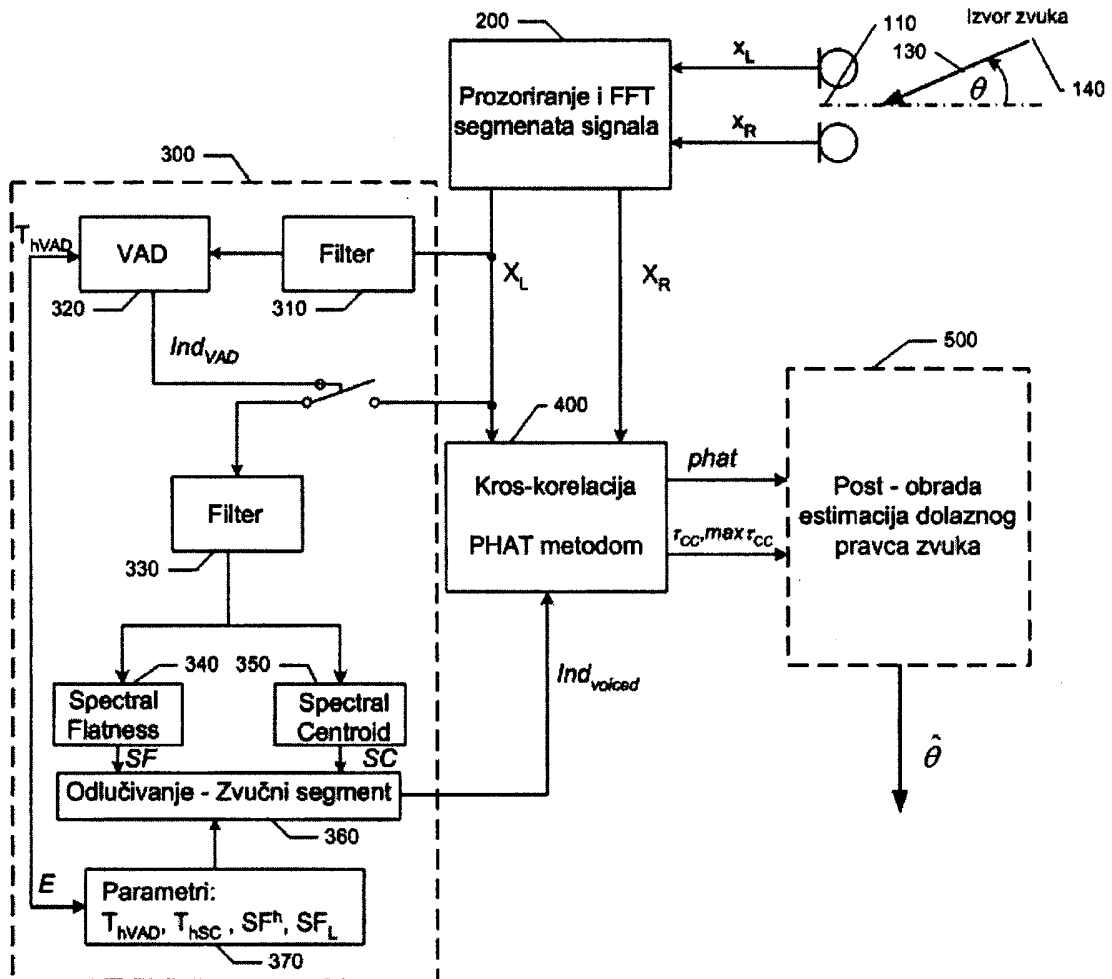
18. Postupak prema zahtevima 15-17 **karakterisan time** što se odvija tako što se računa srednja vrednost  $d$  i varijansa  $s$  poslednjih  $K$  vrednosti ulaznog bafera  $D$  izrazima  $d = \frac{1}{K} \sum_{i=1}^K D(i)$ ,  $s = \frac{1}{K} \sum_{i=1}^K (D(i) - d)^2$  i ako je vrednost varijanse  $s$  ispod prethodno definisanog praga  $Th^{C1}$ , trenutna srednja vrednost  $d$  elemenata bafera  $D$  se dodeljuje najbližem postojećem  $w$ -tom klasteru  $m(w)$ , tj.  $d \in m(w)$ , ako važi  $\|d - m(w)\| < \|d - m(g)\|$ ,  $g=1, \dots, G$ , i  $\|d - m(w)\| < T^{C1}$ , gde je  $\|\cdot\|$  Euklidska norma i gde je prag  $Th^{C1}$  (560) izabran tako da nema preklapanja između klastera, pri čemu ako se sa  $p$  predstavi vektor sa trajanjima svih klastera, onda komponenta vektora  $p$  za dato  $w$  (gde je  $w$  indeks klastera koji je primio  $d$ ) postaje  $p(w) = k$ , a ukoliko je vrednost varijanse  $s$  ipak iznad praga  $Th^{C1}$ , ili nije ispunjen uslov najbližeg klastera  $\|d - m(w)\| < T^{C2}$ , tada se formira novi klaster sa centrom  $m(G) = d$  i izvrši se operacija  $G = G+1$ ,  $G \leq G_{max}$ , gde je  $G_{max}$  maksimalno dozvoljen broj klastera, kao i ako je ispunjen uslov  $(G > G_{max}) \vee (\exists g, (k - p(g)) > T_{max})$  (parametri  $G_{max}$  i  $T_{max}$  su iz 560), tada se "najstariji" klaster uklanja iz skupa klastera, i na kraju se izračunavaju nove varijanse elemenata bafera  $D$  u odnosu na centar svakog klastera  $m(g)$ ,  $g=1, \dots, G$ , tako da vrednost centra klastera za koji je ova vrednost najmanja predstavlja izlaz iz (550) u obliku konačne estimacije dolaznog pravca  $\mathcal{T}_{fin}$ .

19. Postupak prema bilo kojem od prethodnih zahteva **karakterisan time** što se može primeniti na ugao elevacije, odnosno, procenu dolaznog pravca izvora zvuka u vertikalnoj ravni, tako što se prilagodi geometrija mikrofona u vertikalnoj ravni i parametri vezani za međurastojanje mikrofona i prostorni opseg u vertikalnoj ravni.

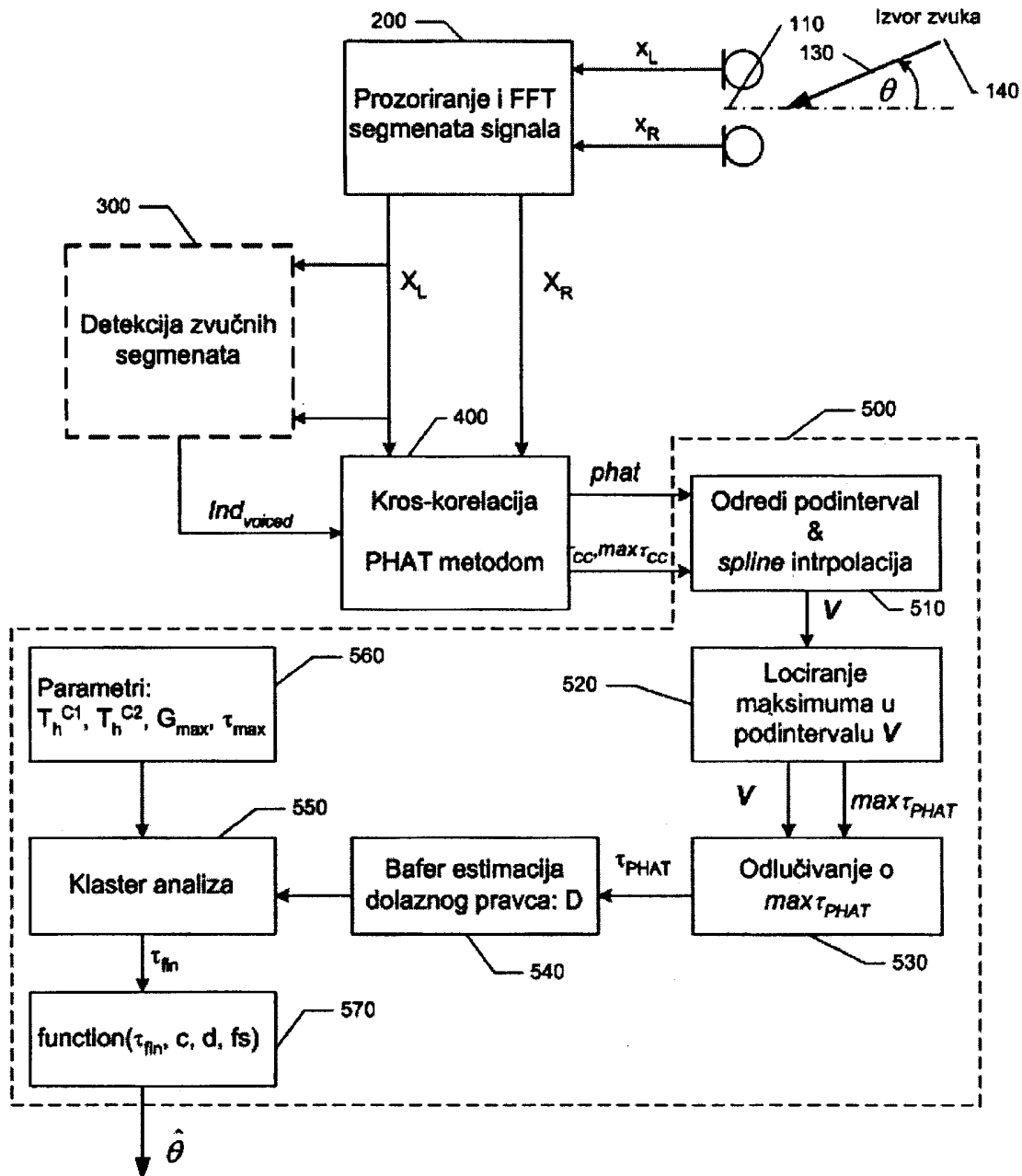




Slika 1.



Slika 2.



Slika 3.