

## INTEGRACIJA TELEFONSKIH FUNKCIJA U TV PRIJEMNIK

Miodrag Temerinac<sup>1</sup>, Nikola Teslić<sup>1</sup>, Zoran M. Šarić<sup>2</sup>, Bojan Mrazovac<sup>2</sup>, Ištvan Papp<sup>2</sup>

<sup>1</sup>Fakultet Tehničkih Nauka, Univerzitet u Novom Sadu, Novi Sad, Srbija

<sup>2</sup>RT-SP d.o.o. signal processing, Novi Sad, Srbija

**Sadržaj:** U radu je predstavljen bežični sistem za dvosmernu govornu komunikaciju integrisan sa postojećom TV tehnologijom. Sistem omogućuje udobniji način telefonske konverzacije korišćenjem mikrofonskog niza i naprednih algoritama za procesiranja glasa, uz istovremenu mogućnost rukovanja uobičajenom TV funkcionalnošću. Obrada signala obuhvata: prostorno filtriranje kojim upravlja modul za otkrivanje pravca i modul za dodatnu obradu, koji uključuje poništavanje zvučnog odjeka i TV zvuka u telefonskoj komunikaciji, uklanjanje stacionarnog šuma i inteligentnu kontrolu jačine zvuka. Svo procesiranje je realizovano u realnom vremenu na platformi zasnovanoj na DSP. Kao komunikacioni kanal mogu se koristiti GSM ili VoIP telefonske veze.

**Ključne reči:** Mikrofonski niz, unapređenje kvaliteta govora, smanjenje šuma, uklanjanje akustičnog šuma, pravac dolaska propagacionog signala, automatsko upravljanje jačinom zvuka, bežična glasovna komunikacija.

### 1. Uvod

Govor je najprirodnije i najrasprostranjenije sredstvo komunikacije između ljudi. Samim tim, značajan broj istraživača usredsređuje svoju pažnju na razvoj tehnologija koje će omogućiti udaljenim osobama da razgovaraju kao da su fizički prisutne jedna pokraj druge. Prvi korak u realizaciji ovog cilja jeste razvoj *Hands-free* komunikacionog sistema. *Hands-free* sistemi rade kao bliski ili udaljeni interfone koji su povezani na telefonsku liniju. U ovom radu se predlaže jedan takav sistem integrisan u TV uređaj koji omogućuje udobniju i jednostavniju komunikaciju u sobnom okruženju.

*Hands-free* komunikacija povlači za sobom niz tehničkih problema kao što su: odjek unutar sobe, akustični eho i različite ambijentalne interferencije. Svaki od navedenih problema se može rešiti na razuman način, međutim, i dalje je prisutan problem njihove integracije u jedinstven sistem koji aktivno reaguje u realnim uslovim. Neki od prethodnih radova na datu tematiku [1] prikazuju rešenje problema unapređenja govora u automobilskom okruženju korišćenjem mikrofonskog niza i adaptivnog prostornog filtra kontrolisanog VAD detektorom. *Hands-free* blok za snimanje [2] koristi adaptivni mikrofonski niz, poništavač odjeka i modul za lokalizaciju izvora zvuka. Navedena rešenja su namenjena sistemima postavljenim na bliskoj udaljenosti.

U ovom radu se opisuje sistem koji obezbeđuje dvosmernu *Hands-free* komunikaciju u složenom akustičnom okruženju. Uobičajeni korisnički scenario je prikazan na slici 1 gde je TV iskorišćen kao komunikaciona platforma. Razvijeni sistem omogućuje obavljanje i prijem poziva međusobno udaljenih učesnika, kao i upravljanje telefonskim imenikom korišćenjem intuitivne grafičke korisničke sprege koja se prikazuje na televizijskom ekranu. Sistem je sa mrežom povezan *Bluetooth* konekcijom. Kao uređaj za izlaz na mrežu koriste se GSM telefon ili računar. S druge strane veze potrebno je aktivirati VoIP aplikaciju na računaru. Glas udaljenog sagovornika se reprodukuje na zvučnicima samog TV uređaja, kombinujući se inteligentno sa zvukom emitovanog televizijskog programa. U slučaju poziva, udaljeni sagovornik prima govor pozivaoca poboljšan naprednim algoritmima za obradu govora. Sistem omogućuje visoko kvalitetnu komunikaciju čak i ukoliko je prisutan zvuk emitovanog televizijskog programa.

U nastavku rada, detaljno su opisani sastavni delovi sistema pojedinačno. U poglavlju 2. sažeto je predstavljena komunikacija i elementi obrade opisanog sistema. U poglavlju 3. detaljno su predstavljeni elementi sistema. Poništavač zvučnog odjeka (*eng. Acoustic echo canceller*), označen kao AEC i prostorno filtriranje su opisani u podpoglavljima 3.1 i 3.2. U podpoglavlju 3.3, autori opisuju realizovano rešenje algoritma lokalizacije izvora zvuka optimizovano za aplikacije u realnom vremenu. Blok za automatsku kontrolu jačine (*eng. Automatic gain control - AGC*) je opisan u podpoglavlju 3.4. Rezultati ispitivanja dobijeni korišćenjem DSP platforme su predstavljeni u poglavlju 4.



Slika 1 *Hands-free* komunikaciona platforma integrisana sa TV uređajem

## 2. Pregled razvojne platforme

Realizovani sistem koristi mikrofonski niz od 5 elemenata da bi efikasno potisnuo izobličenja zvuka. DSP obrađuje signale u realnom vremenu. Razvijeni algoritam potiskuje eho i ambijentalni šum, propuštajući samo signal govora. Poboljšani signal govora se *Bluetooth* konekcijom prenosi do izlaznog uređaja ka mreži (*eng. Gateway*) koji može biti GSM telefon ili računar, te dalje ka udaljenom sagovorniku. DSP i modul za konekciju se nalaze na dodatnoj kartici koja se jednostavno povezuje sa mrežnim kontrolerom. Zadaci mrežnog kontrolera su da usmerava audio kanale na odgovarajući način, da kontroliše modul za konekciju i da obezbeđuje spregu sa korisnikom.

Korišćenje TV uređaja kao *Hands-free* komunikacione platforme indukuje brojne probleme poput nedovoljne jačine govornog signala korisnika koji se nalazi na većoj udaljenosti (najviše do pet metara) od TV uređaja, ali i niza izobličenja kao što su:

odjek koji potiče sa zvučnika TV uređaja, odjek unutar prostorije i različitih stacionarnih i ne-stacionarnih izvora šuma. Oba signala, signal govora pozivanog korisnika i zvuk sa TV uređaja se istovremeno filtriraju višekanalnim akustičnim poništavačem odjeka (*eng. Acoustic echo canceller - AEC*) uz kratko vreme adaptacije i kratko kašnjenje obezbeđeno adaptivnim filtrom poznatijim kao *Partition Block Frequency Domain Adaptive Filter (PBFDAF)* [3]. Referentni signali sa levog i desnog zvučnika se snimaju razdvojenim kanalima AD konvertora s ciljem da učine sistem robustnijim na gubitak paketa prilikom korišćenja VoIP protokola. Višekanalni signal sa mikrofonskog niza je procesiran prostornim filtrom (*eng. Superdirective beamformer*) koji je upravljani modulom za lokalizaciju izvora zvuka (*eng. Sound source localization - SSL*) sposobnim da proceni položaj govornika i u uslovima odjeka. Nepotisnuti stacionarni i ne-stacionarni difuzni šum je dodatno potisnut postfiltrom (PF) zasnovanim na prostornoj različitosti položaja aktivnog govornika i izvora šuma. Modul za automatsku kontrolu jačine zvuka (AGC), na kraju lanca obrade podataka, obezbeđuje jednak nivo jačine i za slabije i za jače govorne signale. AGC je kontrolisan dvostrukim detektorom govora i prostornom informacijom dobijenom sa mikrofonskog niza. Navedena informacija se koristi da spreči pojačanje preostale buke u pauzi govora. Algoritmi su implementirani na DSP platformi *Texas Instruments C6727* i optimizovani za rad u realnom vremenu sa učestanosti odabiranja od 8kHz.

### 3. Pregled modula sistema

#### 3.1 AEC modul

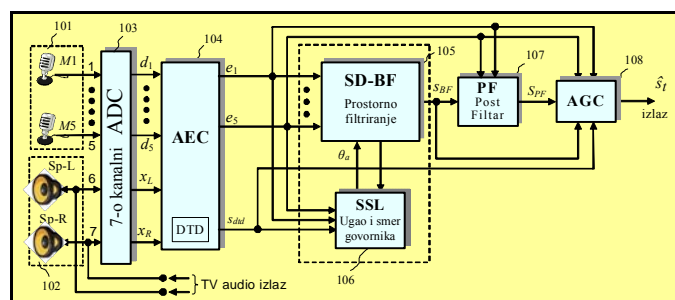
Jedna od osnovnih karakteristika predstavljenog sistema (*eng. Phone TV*), jeste da se korisniku obezbedi nesmetano praćenje TV programa za vreme obavljanja telefonske konverzacije bez potrebe za smanjenjem jačine zvuka na TV uređaju. U ovom slučaju, AEC modul potiskuje govor pozivanog korisnika i zvuk sa TV uređaja. TV zvuk može biti mono ili stereo. Potiskivanje odjeka koji se javlja u stereo zvuku je računski duplo zahtevnije nego u slučaju mono zvuka. U zavisnosti od raspoloživih resursa obrade, sistem može preoblikovati stereo zvuk u mono tokom konverzacije da bi smanjio visoke računске zahteve.

U skladu sa [4] ERLE bi trebao biti veći od 45dB za govor jednog govornika i veći od 30 dB za dijalog. To postiže kombinacijom AEC i prostornog filtriranja. Postfiltrar i AGC modul takođe doprinose potiskivanju preostalog odjeka.

Kombinovano potiskivanje odjeka i prostorno filtriranje se mogu rešiti na dva načina: „AEC upotrebiti prvo“ ili „prostorno filtriranje upotrebiti prvo“ [5]. U strukturi „prostorno filtriranje upotrebiti prvo“, samo filtriranje je u suštini nezavisno od AEC pri čemu se akustični odjek posmatra kao još jedan izvor smetnji. AEC je u tom slučaju realizovan kao jedan adaptivni filter koji je privlačan u pogledu složenosti izračunavanja.

Struktura „AEC upotrebiti prvo“ koristi jedan adaptivni filter za svaki mikروفon. Uz savršeno potiskivanje odjeka, prostorno filtriranje je neometano akustičnim odjekom i koristi sve svoje stepene slobode da potisne prostorno raspodeljene interferencije, kao i odjek unutar prostorije. „AEC upotrebiti prvo“ je upotrebljiv u slučaju da je potrebno koristiti modul za lokalizaciju izvora zvuka (SSL) i postfiltrar (PF). Nedostatak strukture „AEC upotrebiti prvo“ jeste računska zahtevnost, pošto niz od  $M$  mikrofona zahteva približno  $M$  izračunavanja u poređenju sa jednokanalnim AEC.

S ciljem da se obezbede dobre osobine SSL i modula za postfiltriranje, autori koriste „AEC upotrebiti prvo“ strukturu predstavljenu na slici 2. Signali sa mikrofona  $M1$  do  $M5$  se digitalizuju istovremeno sa signalima primljenim sa zvučnika  $Sp-L$  i  $Sp-R$  višekanalnim AD konvertorom (modul 103 na slici 2). Digitalizovani mikrofonski signali  $d1$  do  $d5$  i referentni signali  $xL$  i  $xR$  se dopremaju do AEC modula (104). Izlazi iz AEC modula su signali  $e1$  do  $e5$  sa potisnutim odjekom. AEC modul obuhvata kvalitetan detektor istovremenog razgovora (eng. *Double talk detector* - DTD) radi sprečavanja adaptacije prilikom istovremenog razgovora dva korisnika. Signal detektovanog istovremenog razgovora  $sdt$  je pomoćni izlaz AEC modula koji se koristi u SSL i AGC.



Slika 2 Algoritam obrade glasa

Nekoliko adaptivnih algoritama se mogu koristiti u AEC modulu. Najčešće korišćen algoritam *Time domain normalized least means squares (NLMS)* sadrži kompleksnost koja je linearna u dužini filtra. U uslovima umerenog odjeka unutar prostorije, dužina filtra je nekoliko hiljada tačaka. Računska zahtevnost takvih adaptivnih filtera je previsoka da bi se mogli koristiti u realnom vremenu. Pored toga, navedeni filteri pate od prilično spore konvergencije signala sa obojenim spektrom poput govora. Efikasnija implementacija NLMS algoritma se zasniva na tehnikama frekventnog domena [6]. Adaptivni filter frekventnog domena sprovodi LMS algoritam efikasno pomoću Brze Furijeove Transformacije (eng. *Fast Fourier Transform* - FFT). Pri tom se postiže značajno smanjenje opterećenja sistema za izvršavanje operacija izračunavanja.

Osnovni problem implementacije FLMS algoritma jeste veliko kašnjenje u procesiranju. Na primer, za blok veličine  $N=1024$ , kašnjenje je 128 ms za učestanost odabiranja 8 kHz. Da bi se prevazišao ovaj problem, razvijen je *Partitioned block frequency-domain adaptive filter (PBFDAF)* [3], [7]. Osnovna ideja se ogleda u podeli dugog impulsnog odziva filtra u kraće particije, svaka dužine  $L=N/P$  odbiraka, gde je  $P$  broj particija. Kao rezultat, kašnjenje je umanjeno na  $L$  odbiraka. Radi povećanja stope konvergencije, potrebno je primeniti više iteracija nad istim blokom podataka [3].

Algoritam iskorišćen za realizaciju *Phone TV* sistema je iterativni PBFDAF, ukratko opisan u [3]. Dati algoritam je izmenjen za referentni stereo signal. Autori predlažu korišćenje sedam blokova, pri čemu se svaki sastoji iz 256 tačaka adaptivnog filtra koji obezbeđuje rep od 224 ms. Potiskivanje linearnog odjeka je preko 35 dB za vreme govora samo jednog govornika i preko 30 dB za vreme istovremenog razgovora.

### 3.2 Prostorno filtriranje

Difuzan akustični šum sa jakim refleksijama je u visokoj korelaciji sa direktnim talasom signala lokalnog govornika. U ovom slučaju, tehnike uklanjanja šuma zasnovane

na adaptivnom prostornom filtriranju obezbeđuju slabe rezultate [8], [9], [10]. Stoga se preporučuje korišćenje ne-adaptivnih super direktivnih prostornih filtara (*eng. Super directive beamformer - SDBF*) [11], [12], [13]. Njegov obrazac usmerenosti je optimizovan za idealno difuzno polje šuma po određenim kriterijumima optimizacije [13]. Usmerenost direktivnog prostornog filtra je ka lokalnom govorniku, čiju poziciju određuje modul za lokalizaciju izvora zvuka (SSL – modul 103 na slici 2).

Ulazi super-direkcionog prostornog filtra (SDBF) su signali  $e1$  do  $e5$ . Dati signali su obrađeni u DFT domenu. Signali su razdeljeni na blokove, svaki obuhvata  $L=512$  odbiraka. Preklapanje blokova je 50%. Svaki blok podataka  $\underline{e}_l(n)$ ,  $l=1$  do 5, je izložen odgovarajućoj prozorskoj funkciji  $w$  i transformisan u frekventni domen Brzom Furijeovom Transformacijom (FFT).

$$\underline{E}_l(n) = \mathbf{F}(\mathbf{w}\underline{e}_l(n)) \quad (1)$$

pri čemu  $\underline{E}_l(n) = [E_l(n,1), \dots, E_l(n,L)]^T$  predstavlja vector  $l$ -te kolone DFT skupa  $l$ -tog mikrofonskog kanala  $n$ -tog bloka obrade.  $\underline{e}_l(n) = [e_l(nL/2), \dots, e_l(nL/2 + L - 1)]^T$  je odgovarajući vektor vremenskog domena  $l$ -tog mikrofonskog kanala.  $\mathbf{F}$  označava  $L \times L$  diskretnu Furijeovu transformacionu matricu, dok  $\mathbf{w} = \text{diag}\{w\}$  predstavlja dijagonalnu matricu prozorskog vektora  $w$ . Prozorska funkcija se dodaje dva puta, pre FFT (1) da poboljša rezoluciju učestanosti, a potom i posle obrade u DFT domenu, upravo posle inverzne FFT da spreči klikove na granicama blokova obrade. Ishod je isti ukoliko se primeni prozorska funkcija na mikrofonске signale koja je kvadrat od  $w$ . Prozorska funkcija  $w$  korišćena u priloženom rešenju predstavlja kvadratni koren *Hanning*-ove prozorske funkcije koja omogućuje savršenu rekonstrukciju obrađivanog signala sa 50% preklapanja izlaznih bafera.

Pet mikrofonskih signala grupisanih u vektor  $\mathbf{E}(n,i) = [RE_1(n,i), \dots, E_5(n,i)]$  se obrađuje funkcijom:

$$S_{BF}(n,i) = \mathbf{H}_{sdb}^H(\theta_n, i) \mathbf{E}(n,i) \quad (2)$$

pri čemu  $\mathbf{H}_{sdb}(\theta_n, i)$  predstavlja kolonu vektora mase čiji koeficijenti zavise od ugla upravljanja  $\theta_n$  i metoda super direkcije. DFT koeficijenti poboljšanog zvuka lokalnog govora se preoblikuju u vremenski domen sa funkcijom:

$$\tilde{\underline{s}}_{BF}(k) = \mathbf{w}\mathbf{F}^{-1}([S_{BF}(n,1), \dots, S_{BF}(n,L)]^T) \quad (3)$$

pri čemu je  $\tilde{\underline{s}}_{BF}(n) = [\tilde{s}_{BF}(n,1), \dots, \tilde{s}_{BF}(n,L)]^T$  izlazni bafer odbiraka signala.

Vektor mase  $\mathbf{H}_{sdb}(\theta_n, i)$  se izračunava korišćenjem matrice  $\mathbf{\Gamma}(i)$  idealnog difuznog polja šuma [13] po sledećoj formuli:

$$\mathbf{H}_{sdb}(\theta, i) = \frac{(\mathbf{\Gamma}(i) + \delta\mathbf{I})^{-1} \mathbf{d}_\theta(i)}{\mathbf{d}_\theta^H(i)(\mathbf{\Gamma}(i) + \delta\mathbf{I})^{-1} \mathbf{d}_\theta(i)} \quad (4)$$

Vektor upravljanja  $\mathbf{d}_\theta(i)$  kompenzuje kašnjenje između mikrofona:

$$\mathbf{d}_\theta(i) = [1 \quad e^{-jk_0 d_0 \sin(\theta)} \quad \dots \quad e^{-j(M-1)k_0 d_0 \sin(\theta)}]^T \quad (5)$$

pri čemu  $d_0$  predstavlja rastojanje između susednih mikrofona.

### 3.3 Lokalizacija govornika

Zadatak SSL (*eng. Sound source localization*) modula jeste da odredi poziciju aktivnog govornika čiji će glas biti detektovan mikrofonskim nizom. Pored toga,

utvrđena pozicija govornika se takođe može koristiti za rotiranje kamere u slučaju video konferencija. Postoji nekoliko interferencija koje ometaju lokalizovanje govornika: (a) odjek unutar sobe, (b) ambijentalni šum, (c) zaostali akustični odjek.

Za lociranje izvora zvuka potrebna su najmanje dva mikrofona. Metod ukrštene korelacije se obično primenjuje za procenu vremenske razlike dolaznih signala. Da bi se poboljšala pouzdanost lokalizacije izvora, mnogi od poznatih algoritama koriste informacije koje obezbeđuje nekoliko mikrofona [14], [15]. Značajan broj metoda zasnovanih na estimaciji vremenske razlike (*eng. Time difference estimation - TDE*) osigurava funkcije procene sa više uparenih senzora pre izračunavanja kašnjenja [15], [16], [17]. Drugačiji pristup, označen kao *Steered response power (SRP)* kombinuje snagu usmerenog odgovora sa faznom transformacijom [17], [18].

Primenjeno rešenje generalizuje ukrštenu korelaciju sa faznom transformacijom (GCC-PHAT) adaptiranom za slučaj više mikrofona osiguravajući generalizovanu ukrštenu funkciju korelacije (GCC) različitih mikrofonskih parova pre estimacije vremenske razlike. Predloženi algoritam koristi geometrijske osobine jedinstvenog linearnog mikrofonskog niza za poboljšanje pouzdanosti lokalizacije izvora zvuka i u cilju optimizacije izračunavanja u aplikacijama u realnom vremenu.

Za date mikrofonске signale  $x_i(t)$  i  $x_j(t)$  i njihovu generalizovanu ukrštenu korelaciju  $R_{i,j}(\tau)$ , vremenska razlika dolaska (*eng. Time difference of Arrival - TDOA*) se procenjuje na način opisan u [15].

$$\tau_{i,j} = \arg \max_{\tau \in D} R_{i,j}(\tau) \quad (6)$$

gde se  $R_{i,j}(\tau)$  izračunava iz inverzne diskretne Furijeove transformacije normalizovanog ukrštenog spektra  $G_{i,j}(n)$  po formulama:

$$R_{i,j}(\tau) = \sum_{n=0}^{N-1} G_{i,j}(n) e^{j2\pi n\tau/N}, \quad (7)$$

$$G_{i,j}(n) = \frac{\phi_{i,j}(n)}{\sqrt{E(|X_i(n)|^2)E(|X_j(n)|^2)}}$$

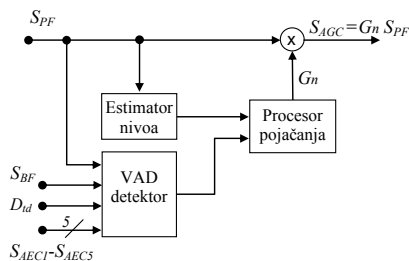
pri čemu  $\phi_{i,j}(n) = E(X_i(n)X_j^*(n))$  predstavlja ukršteni spektar  $n$ -tog DFT odsečka,  $N$  je broj tačaka DFT transformacije,  $X_i(n)$  i  $X_j(n)$  za  $n=0, \dots, N-1$  su DFT transformacije signala sa  $i$ -tog i  $j$ -tog mikrofona ponaosob.  $E()$  je operator očekivanja,  $(\cdot)^*$  označava složeni operator konjugacije. Opseg mogućih TDOA vrednosti je ograničen na konačni interval  $D$ , koji se određuje na osnovu fizičkog razdvajanja između mikrofona.

### 3.4 Automatska kontrola jačine (AGC)

Osnovni zadatak AGC modula je podešavanje jačine govora. Tiši govorni signali treba da budu pojačani, a glasniji utišani. Osnovne osobine AGC modula koji autori predstavljaju su: (a) Dinamički opseg kompresije, (b) Zaštita od pogrešnog povećanja jačine u toku pozadinskog šuma, (c) Zaštita od pogrešnog povećanja jačine prouzrokovana zaostalim odjekom, (d) Klipping kompresije, (e) Detektovanje unutrašnje glasovne aktivnosti zasnovano na više-mikrofonskoj ekstrakciji osobina.

Predstavljeno rešenje AGC modula je inspirisano radom [19] u kom se izračunavanje pojačanja vrši u skladu sa prostornim informacijama izvora zvuka. Blok

dijagram AGC modula je predstavljen na slici 3, sastoji se od estimatora nivoa, VAD detektora i procesora pojačanja koji generiše pojačanje  $G_n$ .

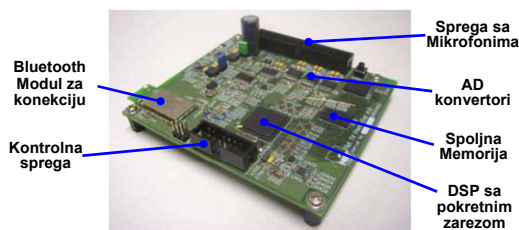


Slika 3 Blok dijagram AGC modula

#### 4. Rezultati ispitivanja

Predstavljeni algoritam audio obrade je testiran u realnom vremenu na DSP platformi. Performanse algoritma su testirane u realnim uslovima korišćenjem nekih objektivnih i subjektivnih mera kvaliteta (ERLE, SNRE, PESQ [20] i test besmislenih slogova). Razumljivost govora je korišćena kao krajnja mera kvaliteta.

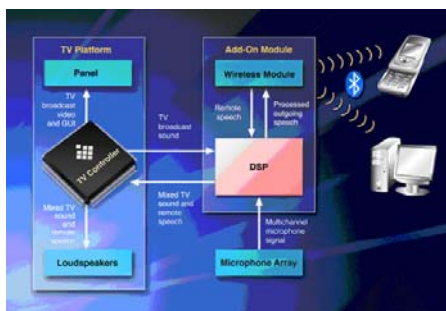
Procena kvaliteta je obavljena implementacijom u realnom vremenu na DSP platformi sa slike 4., koristeći aritmetiku sa pokretnim zarezom na radnoj učestanosti od 300 MHz. Platforma sadrži *Bluetooth* modul za povezivanje, kao i serijsku komandnu spregu za povezivanje sa *Host* TV uređajem. Razvijena platforma je spregnuta sa pet mikrofona korišćenjem AD konvertora i mikrofonskih pojačala. Izlazni signal obrade je prosleđen VoIP ili GSM modulima za izlaz na mrežu (eng. *Gateway*) *Bluetooth* vezom.



Slika 4 Platforma za audio obradu

Kod je razvijen u Matlabu i prilagođen u C kodu sa pokretnim zarezom, pogodnim za implementaciju u realnom vremenu na izabranoj DSP platformi. Za vreme obrade izmereno zauzeće DSP je bilo 80% uključujući obradu signala i zauzeće pristiglo iz DSP okruženja. Zauzeće memorije je bilo 512 kB.

Štampana ploča za audio obradu je integrisana u sam TV uređaj kao dodatni modul (Slika 5). Mikrofonski niz je ugrađen u ram TV uređaja (Slika 6 - levo). Standardna TV programska podrška je proširena programskom podrškom za komunikaciju i kontrolu DSP štampane ploče. Na ovaj način je TV uređaj postao komunikacioni uređaj pogodan za interaktivnu *Hands-free* komunikaciju.



Slika 5 Integracija audio obrade u TV uređaj

Testiranje je obavljano u okruženju prikazanom na slici 6 - desno (prostorija 1). Dimenzije prostorije su 7m x 5m x 2.6m, sa zapreminom 90m<sup>3</sup>. Prostorija predstavlja uobičajenu kancelariju sa ravnim zidovima. Odjek unutar prostorije je iznosio 400 ms.

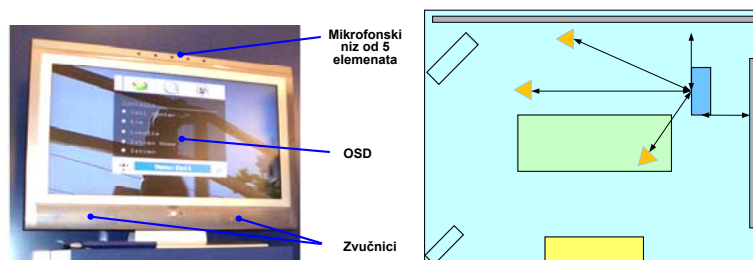
U nizu testova, sledeći testni slučajevi su bili varijabilni: (a) Jedan korisnik govori ili istovremeno oba korisnika govore (s jedne i druge strane komunikacione veze), (b) Jačina zvuka pozivaoca, (c) Položaj pozivaoca (od 0 do 30 stepeni), (d) Različite vrste signala govora pozivanog govornika (govor ili muzika: pop, rok, ambijentalna), (e) Različite vrste šuma (razgovor, bučno okruženje, slavlja, muzika, nestacionarni šum).

Kompletno testiranje je rezultovalo sa 22 različita testna slučaja koji pokrivaju očekivane okolnosti korišćenja u svakodnevnom životu. Da bi se obezbedila valjana reprodukcija, nivoi signala su prilagođeni korišćenjem merača nivoa zvuka. Signali koji odgovaraju testnim slučajevima su snimljeni. Objektivne mere ERLE, SNRE i PESQ su preračunate van relanog vremena korišćenjem snimljenih signala i prikazani na slici 7.

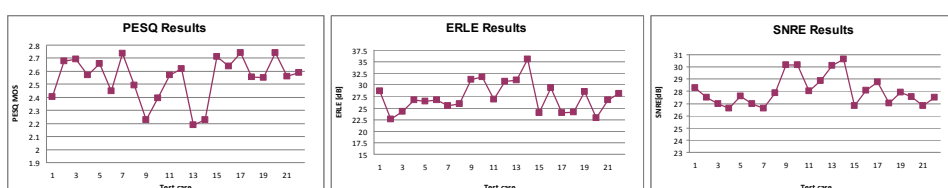
Ista prostorija je korišćena kao okruženje za ispitivanje razumljivosti govora. Besmisleni slogovi su reprodukovani kao signali govornika, zajedno sa izobličenjima – odjekom i šumom. Testna osoba se nalazila u zvučno izolovanoj prostoriji (prostorija 2). Obraden glas je prenet u prostoriju 2 korišćenjem VoIP veze i reprodukovan na zvučnicima. Rezultati prepoznavanja besmislenih slogova predstavljeni su na slici 8. Pored toga, sistem koji omogućuje 75% uspešno prepoznatih besmislenih slogova, istovremeno obezbeđuje 100% prepoznavanje reči iz normalnog govora [21].

Dobijeni su sledeći rezultati: prosečna PESQ vrednost iznosi oko 2.6 u oba slučaja i kad lokalni govornik priča, kao i u slučaju istovremenog razgovora oba govornika, i lokalnog (pozivalac) i udaljenog (pozivani). Prosečna prigušenost odjeka (ERLE) duž skupa testnih vektora iznosi 27 dB. Prosečna prigušenost šuma (SNRE) uključujući nestacionaran šum, iznosi 28 dB.





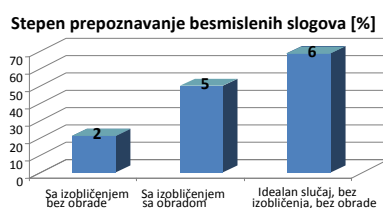
Slika 6 Uređaj razvijen za testiranje u realnom vremenu (levo) i Testno okruženje (desno)



Slika 7 PESQ, ERLE i SNRE pregled rezultata

## 5. Zaključak

Razvijeni sistem obezbeđuje dvosmernu glasovnu komunikaciju visokog kvaliteta u *Hands-free* režimu rada. Ciljni uslovi korišćenja su u okruženjima poput dnevnog boravka ili kancelarije. Razvijeni audio podsistem je integrisan sa TV uređajem. Rešenje je primenljivo u obliku *Hands-free* komunikacionog terminala sa naprednim osobinama poput istovremenog korišćenja TV funkcionalnosti, dvosmerne komunikacije i visokog kvaliteta glasa prenetog na daljinu. Razvijeni sistem omogućuje konformnu konverzaciju korišćenjem neke od komunikacionih tehnologija (GSM ili VoIP). Razvijena tehnologija se može iskoristiti u sistemima poput opreme za bezbedno telefoniranje u automobilima, telekonferencijskih sistema i sistema zasnovanih na glasovnoj sprezi ljudi i mašina.



Slika 8 Rezultati prepoznavanja besmislenih slogova

## Literatura

- [1] J.-S. Hu, C.-C. Cheng, W.-H. Liu, and C.-H. Yang, "A Robust Adaptive Speech Enhancement System for Vehicular Applications," *IEEE Trans. on Consumer Electronics*, vol.52, no.3, pp.1069-1077, Aug.'06.
- [2] K. Kobayashi, Y. Haneda, K. Furuya, and A. Kataoka, "A Hands-Free Unit with Noise Reduction by Using Adaptive Beamformer," *IEEE Trans. on Consumer Electronics*, vol.54, no.1, pp.116-122, Feb.'08.

- [3] K. Eneman, M. Moonen, "Iterated Partitioned Block Frequency-Domain Adaptive Filtering for Acoustic Echo Cancellation," *IEEE Trans. on Speech and Audio Processing*, vol.11, no.2, pp.143-158, Mar.'03.
- [4] ITU-T recommendation G.167, *Acoustic echo controllers*, Mar. 1993.
- [5] W. L. Kellermann, "Acoustic Echo Cancellation for Beamforming Microphone Arrays," in M. Brandstein and D. Ward (eds), *Microphone Arrays*, New York: Springer, 2001, pp.281–306,
- [6] Earl R. Ferrara, "Fast Implementation of LMS Adaptive Filters," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 4, pp. 474-475, Aug. 1980,
- [7] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 373-376, Feb. 1990.
- [8] Z. M. Saric, and S.T. Jovicic, "Adaptive microphone array based on pause detection", *Acoustics Research Letters Online (ARLO) 5(2)*, pp. 68-74, Apr. 2004.
- [9] S.T. Jovicic, Z.M. Saric, S.R. Turajlic, "Application of the maximum signal to interference ratio criterion to the adaptive microphone array," *Acoustics Research Letters Online (ARLO) 6(4)*, pp.232-237, Oct.'05.
- [10] I. Papp, Z. Saric, S. Jovicic, N. Teslic, "Adaptive microphone array for unknown desired speaker's transfer function", *Journal of Acoustic Society of America, Express Letters*, pp. 44-49, July 2007.
- [11] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques" in M. Brandstein and D. Ward (eds), *Microphone Arrays*, New York: Springer, 2001, pp. 36–60.
- [12] C.Marro, Y.Mahieux, K.U.Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Process.*,vol.6,no.3, May'98
- [13] J. Bitzer, and K. U. Simmer, "Superdirective microphone arrays," in M. Brandstein and D. Ward (eds), *Microphone Arrays*, New York: Springer, 2001, pp. 19-38.
- [14] K. Kwak, and S. Kim, "Sound source localization with the aid of excitation source information in home robot environments," *IEEE Trans. Consumer Electronics*, vol. 54, no. 2, pp. 852-856, May 2008.
- [15] T. Nishiura, T. Yamada, S. Nakamura, and K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," in *Proc. of IEEE ICASSP'00*, vol.2, pp.1053–1055, Istanbul, Turkey, June 2000
- [16] S. M. Griebel and M. S. Brandstein, "Microphone array source localization using realizable delay vectors," in *Proceedings of IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA '01)*, pp. 71–74, New Platz, NY, USA.
- [17] J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on applied signal processing*, vol. 2006, pp. 1–19, 2006.
- [18] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in M. Brandstein and D. Ward (eds), *Microphone Arrays*, New York: Springer, 2001, pp. 157–180.
- [19] K. Kobayashi, Y. Haneda, K. Furuya, and A. Kataoka, "A hands-free unit with adaptive microphone array for directional AGC," in *Acoustic echo and noise control : ninth international workshop, IWAENC 2005*, September 12-15, 2005, Eindhoven, The Netherlands,
- [20] ITU-T, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", Inter. Telecom. Union, 2001.
- [21] H. Levitt, J.C. Webster, "Effects of Noise and Reverberation on Speech", In C.M. Harris (eds), *Handbook of Acoustical Measurements and Noise Control*, Chapter 16, McGraw-Hill, 1991.

**Abstract:** *This paper presents a system for full-duplex hands free voice communication integrated with TV technology. The system provides comfort conversation by utilization of microphone array and advanced voice processing algorithms, even with simultaneous TV usage. Signal processing includes superdirective beamformer steered by direction-finding module, postprocessing module, acoustic echo canceller, stationary noise reduction module and automatic gain control. All processing is realized in real-time on DSP based platform. As communication channel GSM or VoIP can be used.*

**Keywords:** *Microphone array, speech enhancement, noise reduction, acoustic noise cancellation, direction of arrival, automatic gain control, hands-free communications*

#### HANDS-FREE VOICE COMMUNICATION OVER TV

Miodrag Temerinac, Nikola Teslic, Zoran M. Saric, Bojan Mrazovac, Istvan Papp