Poboljšanje kvaliteta govornog signala u realnom okruženju MMB-SS algoritmom

Z. Veličković, *Visoka tehnička škola strukovnih studija, Niš*

***Sadržaj* — U ovom radu su razmatrane performanse predloženog mel-multi-band (MMB) SS algoritma na rekonstrukciju govora kod DSR sistema u uslovima realnih smetnji. Obzirom da smetnje iz realnog okruženja imaju vremenski promenljiv frekvencijski spektar, njihov uticaj na govorni signal je redukovan u skladu sa psihoakustičkim modelom baziranom na MEL skali. Objektivnim metodom određene su performanse MMB SS algoritma za tipične smetnje iz realnog okruženja i pokazana je superiornost nad standardnim SS algoritmom.**

***Ključne reči* — Spektralno oduzimanje, DSR sistemi, MEL filterska banka, Rekonstrukcija govora.**

I. UVOD

MOBILNE govorne aplikacije zahtevaju pouzdanu komunikaciju u svim realnim okruženjima. U zavisnosti od sredine, na signal govora se mogu superponirati različite vrste smetnji iz realnog okruženja. U prisustvu ovih smetnji performanse mobilnih govornih aplikacija značajno se degradiraju. Od izuzetne važnosti je unaprediti perceptualni aspekt govora a posebno njegovu razumljivost. Problem izdvajanja čistog govora iz govornog signala kontaminiranog realnim smetnjama je razmatran dugi niz godina. Značaj ovog problema je posebno aktuelizovan kod aplikacija u sistemima za automatsko prepoznavanje govora (engl. *Automatized Speech Recognition* - ASR). Pored ASR sistema koji su bazirani na aplikovanom kodeku govornog signala, razvijeni su distribuirani sistemi za prepoznavanje govora (engl. *Distributed Speech Recognition* - DSR). Kod DSR sistema [1] parametri prepoznavanja se izračunavaju na terminalnoj strani, tako što se govorni signal deli na frejmove (obično 25ms) za koje se određuju parametri prepoznavanja. Parametri prepoznavanja se pakuju u RFV (engl. *Recognition Feature Vector* - RFV) i šalju serverskoj strani na prepoznavanje. Savremeni servisi mobilnih aplikacija pored prepoznavanja govora, često zahtevaju i rekonstrukciju govornog signala iz podataka za prepoznavanje, odnosno RFV-a. Algoritmi ekstrakcije RFV-a se može se naći u radu [2]. U [3] je razmatran uticaj dužine MFCC vektora na kvalitet rekonstruisanog govora iz MFCC koeficijenata u prisustvu belog Gausovog šuma. Kada se DSR sistemi primenjuju u okruženju u kome je prisutan aditivni šum, dolazi do značajne degradacije performansi sistema, a samim tim i kvaliteta rekonstruisanog govora. Potiskivanje šuma govornog signala zasnovano na spektralnom oduzimanju (engl. *Spectral Subtraction* - *SS*) je jedan od prvih algoritama koji je dao značajne rezultate u prisustvu aditivnih smetnji [4]. SS algoritam je posebno našao primenu u situacijama kada je na raspolaganju samo signal iz jednog mikrofona. Ovaj scenario odgovara primeni u mobilnim govornim komunikacijama. Zahvaljujući jednostavnoj aplikaciji i značajnoj redukciji signala šuma, ovo je najčešće korišćen algoritam za potiskivanje šuma kod govornih servisa. Koncept spektralnog oduzimanja se bazira na razlici nekorelisanih spektara snage kontaminiranog govornog signala i snage šuma. Nepoznati spektar snage šuma se može izračunati iz negovornog dela sekvence. Međutim, obzirom da izračunati spektar šuma ne odgovara uvek stvarnoj vrednosti, pri oduzimanju spektara, generišu se spektralne komponente koje će proizvesti karakterističan muzički šum. U cilju snižavanja muzičkog šuma redukuju se ekstremi u rezultujućem spektru snage. Jedna modifikacija SS algoritama je prikazana u [5]. Modifikacija baznog SS algoritma se odnosi na oduzimanje nelinearne funkcije spektra šuma uvođenjem „*over-subtraction*“ faktora α . Poznato je da vrednost SNR-a tokom trajanja govora u realnom okruženju nije konstantna. Da bi se ova činjenica uzela u obzir, favorizuje se dinamičko izračunavanje „*over-subtraction*“ faktora α u funkciji lokalnog SNR-a. Sa druge strane, važna osobina realnih smetnji je da poseduju neuniformni frekvencijski spektar. Potiskivanje ovih smetnji modifikovanim SS algoritmom, je realizovano zahvaljujući adaptivnom „over-subtraction“ faktoru. Istraživački timovi [4] razmatraju uticaj reda generalizovanog SS metoda. Pokazano je da dinamička promena reda SS modela može unaprediti potiskivanje šuma.

U ovom radu je predloženo potiskivanje realnih smetnji govornog signala kod DSR sistema. Predloženo je da se potiskivanje ovih smetnji obavlja u fazi predobrade signala govora primenom SS algoritma pre određivanja RFV-a. U radnom okruženju gde su prisutne smetnje iz realnog života predloženo je dinamičko prilagođene primenom „over-subtraction“ faktora α . Deljenjem frekvencijskog spektra na segmente i pridruživanjem vrednosti parametra α , obezbeđuje se prilagođenje SS algoritma frekvencijskom spektru realnih smetnji. Frekvencijski segmenti za koje se izračunava vrednost parametra α su nejednake dužine, i određeni su u skladu

Z. Veličković, Visoka tehnička škola strukovnih studija Niš, Aleksandra Medvedeva 20, Srbija (telefon:+381-18-588-211; faks: 381-18-588-211; e-mail: zoran.velickovic@vtsnis.edu.rs).

sa psihoakustičkim modelom čovečije percepcije zvuka. Ovo je ključna razlika koja izdvaja ovaj rad u odnosu na algoritme prikazane u [5]. U nastavku rada u poglavlju 2 je dat prikaz baznog SS algoritma. U poglavlju 3 data je predložena modifikacija baznog SS algoritma. Modifikovani SS algoritam u ovom radu referenciramo sa MMB SS. U sekciji 4 su prezentirani eksperimentalni rezultati za različite tipove smetnji iz realnog života, a zaključna razmatranja su data u poglavlju 5.

II. SPEKTRALNO ODUZIMANJE

A. Bazni algoritam spektralnog oduzimanja

Ako se čistom govornom signalu $s(n)$ superponira šum $d(n)$ ima se degradirani signal govora $x(n)$:

$$x(n) = s(n) + d(n), \quad n = 0,1,2,..., N-1, \tag{1}$$

gde N predstavlja dužinu govorne sekvence. Spektar snage degradiranog govora se može izračunati:

$$|X(k)|^2 = |S(k)|^2 + |D(k)|^2 + S(k)\cdot D^*(k) + D(k)\cdot S^*(k), \tag{2}$$

gde * predstavlja operator kompleksne konjugacije. Funkcija $|S(k)|^2$ je spektar snage govornog signala, a $|D(k)|^2$ spektar snage šuma. Pod pretpostavkom da su signal govora i signal smetnji nekorelisani, članovi zbira u (2) koji sadrže cross-spectrum $S(k)\cdot D^*(k)$ i $D(k)\cdot S^*(k)$ mogu se zanemariti [5], tako da se iz (2) dobija:

$$|\hat{S}(k)|^2 \approx |X(k)|^2 - |D(k)|^2 . \tag{3}$$

Pre izračunavanja spektra degradiranog signala, na degradirani govorni signal primenjuje se Hammingov prozor radi segmentiranja govornog signala u frejmove. Frekvencijski spektar frejma degradiranog govornog signala je određen diskretnom Furieovom transformacijom (engl. *Discrete Fourier Transform* - DFT):

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi k n}{N}} = |X(k)| e^{j\varphi_x(k)} \tag{4}$$

gde $|X(k)|$ predstavlja moduo a $\varphi_x(k)$ fazu spektra degradiranog govornog signala i $0 \le k \le N-1$.

Spektar snage šuma $|D(k)|^2$ se ne može direktno izračunati, te se koristi sledeća aproksimacija:

$$|D(k)|^2 \approx E\left[|D(k)|^2\right] = |\hat{D}(k)|^2, \tag{5}$$

gde $E[\bullet]$ predstavlja operator matematičkog očekivanja i izračunava se za frejmove u kojima nije prisutan govor. Zamenom (5) u (3) dobija se:

$$|\hat{S}(k)|^2 \approx |X(k)|^2 - |\hat{D}(k)|^2 . \tag{6}$$

B. Algoritam nelinearnog spektralnog oduzimanja

Modifikovan spektar snage govornog signala se može izračunati:

$$|\hat{S}(k)|^2 \approx |X(k)|^2 - \alpha |\hat{D}(k)|^2 , \ \alpha \ge 1, \tag{7}$$

gde je α „*over-subtraction*“ faktor. „Over-subtraction“ faktor α minimizira ostatak tako što usrednjenu vrednost snage šuma pomnoženu „over-subtraction“ faktorom oduzima od spektra degradiranog govornog signala. Da bi se sprečila situacija u kojoj bi rezultujući spektar bio negativan, primenjuje se sledeća popravka modifikovanog spektra snage:

$$|\hat{S}(k)|^2 = \begin{cases} |\hat{S}(k)|^2, & |\hat{S}(k)|^2 > \beta |\hat{D}(k)|^2 \\ \beta |\hat{D}(k)|^2, & drugde \end{cases} . \tag{8}$$

gde je β „*spectral floor*“ parametar. Ova popravka dovodi do toga da se negativne vrednosti rezultujućeg spektra dovode na nivo niži od nivoa srednjeg šuma ($\beta << 1$). Modifikovani spektar govora se dobija korišćenjem modua popravljenog spektra $|\hat{S}(k)|$ i faze degradiranog signala $\varphi(k)$na sledeći način:

$$\hat{S}(k) = |\hat{S}(k)| e^{j\varphi_x(k)}. \tag{9}$$

SS metod može biti generalizovan na sledeći način:

$$|X(k)|^\gamma \approx |S(k)|^\gamma + \alpha |D(k)|^\gamma , \tag{10}$$

gde se za vrednost parametra $\gamma = 1$ ima amplitudski spektar, dok se za za $\gamma = 2$ ima spekar snage.

C. Algoritam multi-band spektralnog oduzimanja

U realnom okruženju frekvencijski spektar nije uniforman u celom govornom opsegu. Posledica neuniformne frekvencijske raspodele kod realnih smetnji je da faktor α nije uniforman u celokupnom frekvencijskom opsegu. Zbog toga, u pojedinim govornim frejmovima primena SS metoda može rezultirati muzičkim šumom. Da bi se izračunavanje parametra α prilagodilo realnim uslovima, u [5] se preporučuje podela frekvencijskog spektra šuma i govora u više jednakih segmenata. Za svaki frekvencijski segment *i* izračunava se segmentirani „over-subtracted“ faktor α_i. Ako se govorni spektar podeli na N neprekloplјenih segmenata čist govor se može izračunati na sledeći način:

$$|\hat{S}_i(k)|^2 \approx |X_i(k)|^2 - \alpha_i |\hat{D}_i(k)|^2 , \ w_i < k < w_{i+1}, \tag{11}$$

gde k predstavlja indeks DFT koeficijenta. Koeficijenti w_k i w_{k+1} predstavljaju početni i krajnji DFT koeficijent *i*-tog segmenta. Za svaki frekvencijski segment *i* izračunava se „over-subtracted“ faktor α_i koji zavisi od vrednosti SNR-a u posmatranom segmentu $NSNR_i$:

$$NSNR_i = 10\log_{10}\left[\sum_{k=w_i}^{w_{i+1}}\left|X_i(k)\right|^2 \Big/ \sum_{k=w_i}^{w_{i+1}}\left|\hat{D}_i(k)\right|^2\right]. \quad (12)$$

Segmentni „over-subtraction“ faktor α_i se određuje:

$$\alpha_i = \begin{cases} 1, & NSNR_i \geq 20dB \\ \alpha_0 - \frac{3}{20}\cdot NSNR_i, & -6dB \leq NSNR_i \leq 20dB \\ 4.9, & NSNR_i \leq -6dB \end{cases}, \quad (13)$$

gde je $\alpha_0 = 4$ željena veličina over-subtracted faktora na 0dB SNR-a. Negativne vrednosti rezultujućeg spektra se snižavaju i postavljaju na predefinisanu vrednost određenu parametrom β (8).

III. MEL-MULTI-BAND (MMB) SS ALGORITAM

Čovekova percepcija zvuka nije linearna funkcija frekvencije [1]. Razvijeno je nekoliko nelinearnih skala za opis ove zavisnosti [6]. Kod DSR sistema koriste se MEL-filtarske banke za opis auditivnog sistema čoveka. Familije MEL-filtarskih banaka se razlikuju po broju i tipu primenjenih filtara. Tako, filtarska banka FB_SLANEY sadrži 40 filtara, dok filtarska banka FB_MEL sadrži 23 filtra [6]. MEL-filtri poseduju trougaone spektralne karakteristike sa različitim širinama propusnih opsega. Istovremeno, centralne frekvencije trougaonih filtara f_c su nelinearno razmaknute na frekvencijskoj osi. Centralne frekvencije trougaonih filtara iz FB_MEL filtarske banke prema ETSI Aurora [1] standardu se određuju:

$$f_c(m) = Mel^{-1}\left\{Mel(f_{start}) + \frac{Mel(f_s/2) - Mel(f_{start})}{K+1}\cdot m\right\},$$
$$m = 1,2,...,M, \quad (14)$$

gde je *K=23* ukupan broj filtara, a *m* redni broj filtra u filtarskoj banci, dok je MEL skala definisana:

$$Mel(x) = 2595\cdot\log_{10}\left(1 + \frac{x}{700}\right). \quad (15)$$

Primer realizacije FB_MEL filtarske banke može se naći u [6]. MEL filtri sa većom centralnom frekvencijom f_c, imaju veći propusni opseg, odnosno, filtri sa manjom centralnom frekvencijom imaju manji propusni opseg. Ovako definisani propusni opsezi predstavljaju frekvencijske segmente, MMB segmente, za realizaciju MMB SS algoritma. Segmentni SNR ($NSNR_i$) se izračunava korišćenjem spektralnih komponenata za svaki segment:

$$\sum_{k=w_i}^{w_{i+1}}\left|X_i(k)\right|^2, \sum_{k=w_i}^{w_{i+1}}\left|\hat{D}_i(k)\right|^2 \quad (16)$$

Donja i gornja granica formiranih MMB segmenata je definisana sa w_i i w_{i+1} respektivno.

IV. EKSPERIMENTALNI REZULTATI

A. Metrika kvaliteta rekonstruisanog govora

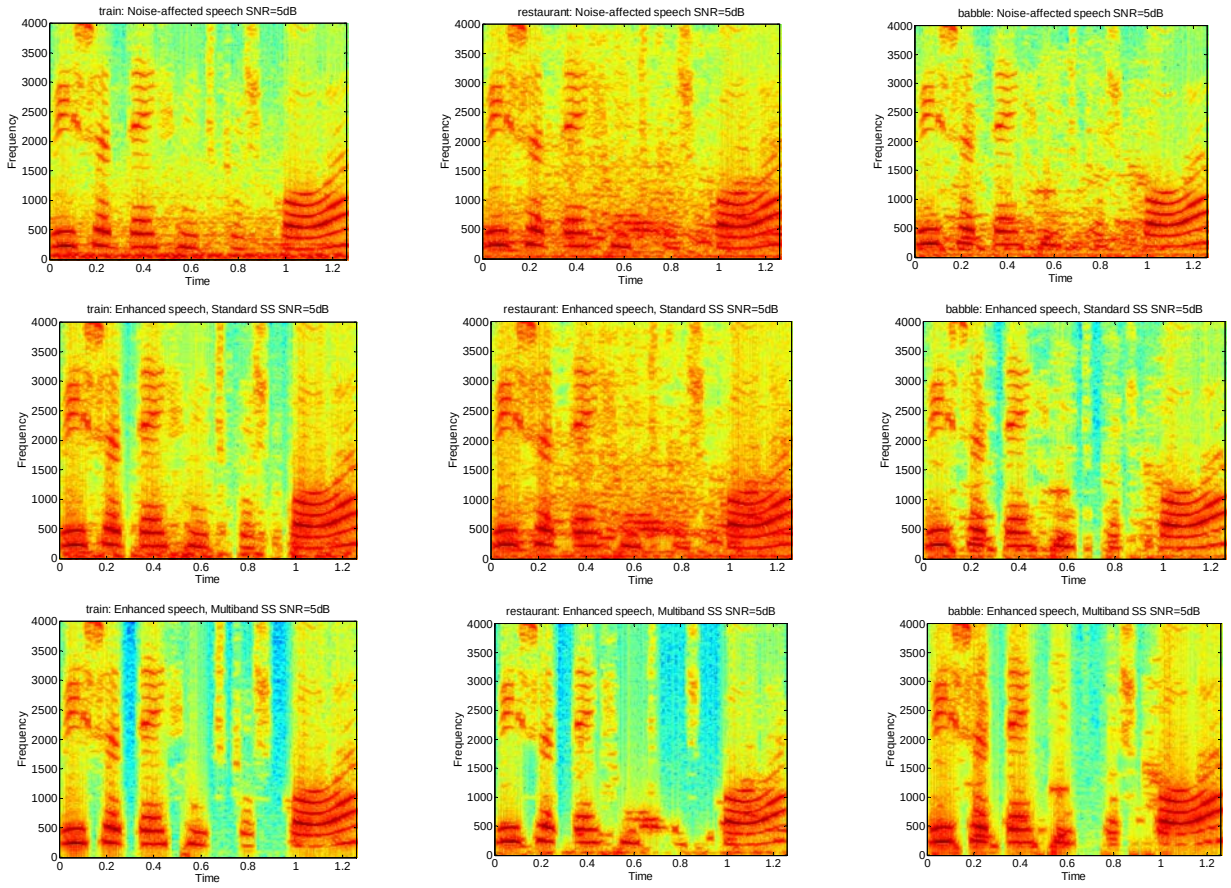
U ovom radu su određene performanse MMB SS algoritma implementiranog u fazi predobrade govornog signala kod DSR sistema. Za upoređivanje kvaliteta rekonstruisanog govornog signala neophodno je uspostaviti odgovarajuću metriku. Kvalitet rekonstruisanog govora određen je procenom percepualnog audio kvaliteta (engl. *Perceptual Evaluation of Audio Quality* - PEAQ). ITU-T preporukom P.862 preporučuje se korišćenje PESQ-a, međutim, korišćen je PEAQ jer nije raspoloživa pouzdana verzija PESQ-a za Matlab. Vrednosti ODG-a (engl. *Objective Difference Grade*) predstavlja ocenu kvaliteta rekonstruisanog govornog signala. Vrednosti ODG ocene kreću se u rasponu od -4 do 0. Ocena ODG = 0 predstavlja neprimetnu degradaciju audio kvaliteta, dok ocena ODG = -4 predstavlja neprijatnu distorziju audio kvaliteta. Performanse MMB-SS algoritma su određene za različite vrednosti SNR-a i tipične smetnje iz realnog života. Spektrogrami originalnog i rekonstruisanog govora su takođe značajni objektivni pokazatelji kvaliteta rekonstrukcije govora. Spektrogram pokazuje promenu frekvencijskog spektra govornog frejma u vremenu, te se uvidom u spektrograme originalnog i rekonstruisanog govornog signala može dati procena kvaliteta rekonstrukcije.

B. Baza govornih signala

Baza govornih signala se formira uzorkovanjem govornog signal iz mikrofona frekvencijom F_s=*8kHz* i njegovim arhiviranjem [3]. Baza realnih smetnji je raspoloživa na URL adresi http://www.utdallas.edu/~loizou/speech. Proširivanjem baze govornih signala, novoformiranim signalima dobijenim superponiranjem skalirane vrednosti realnih smetnji, formirana je baza signala za određivanje performansi MMB-SS algoritma. Na ovaj način dobijena je *proširena baza govornih signala* sa definisanim SNR odnosom [3].

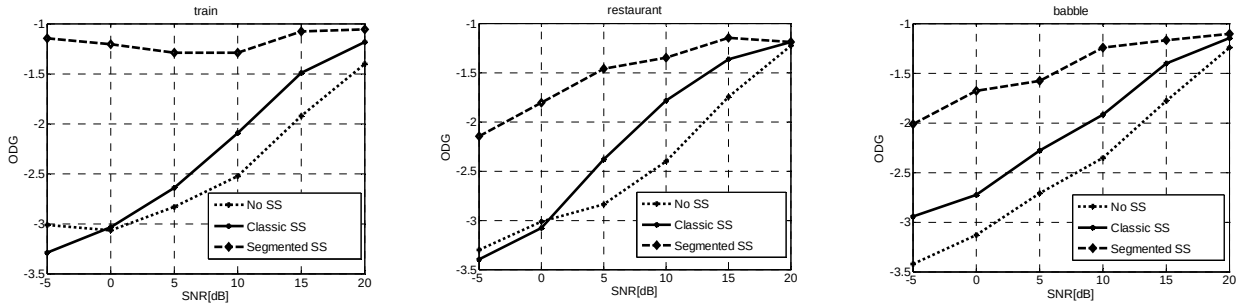
C. Analiza dobijenih rezultata

Na slici 1 prikazani su spektrogrami dobijeni iz rekonstruisanih govornih signala za karakteristične smetnje snimljene u vozu, restoranu i pri istovremenom govoru više govornika (train, restourant i babble) za SNR=5dB. Uklanjanje pojedinih spektralnih delova smetnji iz rekonstruisanih govornih signala je uočljivo za oba primenjena algoritma: Standard SS i MEL Multiband SS. Sa spektrograma na slici 1a, 1b i 1c uočava se značajno uklanjanje spektra smetnji iz rekonstruisanih signala. Na slici 2 su prikazane dobijene ODG ocene primenom PEAQ algoritma. Grafici predstavljaju ODG ocene u funkciji SNR-a za tri komparirana algoritma. Naime, prvi algoritam, No SS, zapravo ne primenjuje ni jednu tehniku za redukciju smetnji iz realnog okruženja. Drugi algoritam je standardni SS algoritam (classic SS), dok je treći preporučen MMB-SS algoritam (Segmented SS). Sa svih grafika se jasno može uočiti da sa porastom SNR-a svi algoritmi zaslužuju veće ocene. Drugim rečima, audio percepcija kvaliteta rekonstruisanog govora raste. Porast ODG ocene nije jednak za sve tipove realnih smetnji. Najveće poboljšanje se dobija za smetnje tipa „train“. Kod svih slika na vrhu se nalazi grafik koji pripada preporučenom MMB-SS algoritmu, čime se potvrđuje njegova superiornost u odnosu na klasični SS algoritam.



a) b) c)

Sl. 1. Spektrogrami originalnog govornog signala kontaminiranog a) "train", b) "restaurant" i c) "babble" smetnjom sa SNR=5dB (najviši red). Spektrogrami rekonstruisanih govornih signala primenom standardnog SS algoritma (srednji red) i MMB SS algoritma (najniži red).



a) b) c)

Sl. 2. Perceptualna procena audio kvaliteta (PEAQ) u funkciji SNR-a za govorni signal kontaminiran a) "train", b) "restaurant" i c) "babble" smetnjama za originalni signal (No SS), signal kod koga je potisnut šum standardnim SS algoritmom (Classic SS), odnosno MMB SS algoritmom (Segmented SS).

LITERATURA

[1] ETSI, "ES 202 050, v1.1.5, STQ, DSR, Advanced front-end feature extraction algorithm compression algorithms", 2007.

[2] Zoran Milivojević, Zoran Veličković, *"Performances of MFCC algorithm in the presence of the white Gaussian noise "*,*UNITECH 08*, Gabrovo, pp. I-190- 195, 2008.

[3] Z. Veličković, Z. Milivojević, "Popravka kvaliteta govornog signala sa superponiranim belim gausovim šumom ss algoritmom kod DSR sistema", ETRAN 09.

[4] Dan Ellis, "Speech & Audio Processing & Recognition", http://www.ee.columbia.edu/~dpwe/e6820/, 2006.

[5] R.M. Udrea, N. Vizireanu, S. Ciochina S. Halunga "Nonlinear spectral subtraction method for colored noise reduction using multi-band Bark scale", *Signal Processing 88*, pp. 2764– 2776, 2008.

[6] Z. Veličković, Z. Milivojević, *"Performanse MEL filtarskih banaka kod rekonstrukcije govora*", Informacione tehnologije - IT 2009.

ABSTRACT

The performances of proposed mel-multi-band (MMB) SS algorithm on speech reconstruction in DSR systems in real-life noise are considered. The speech noise reduction is applied based on mel scale, and superiority of MMB-SS algorithm in real-life environment is presented.

MMB-SS ALGORITHM FOR ENHANCING SPEECH IN REAL-LIFE ENVIRONMENT

Zoran Veličković