

REPREZENTOVANJE GOVORNOG SIGNALA ALGORITMOM SA SIGNATURE FUNKCIJAMA

Zoran Milivojević, *Viša tehnička škola, Niš*
 Milorad Mirković, *IRITEL, Beograd*
 Predrag Rajković, *Mašinski fakultet, Niš*
 Borivoje Milošević, *Viša tehnička škola, Niš*

Sadržaj – U prvom delu rada opisan je algoritam za reprezentovanje govornog signala koji je baziran na signature (potpisnim) funkcijama. Potpisne funkcije generisane su na bazi statističkih parametara govornog signala. Analiziran je efekat algoritma posredstvom faktora kompresije. U drugom delu rada prikazani su rezultati primene opisanog algoritma kod reprezentovanja govornog signala srpskog jezika. Rezultati su dobijeni posredstvom MOS testa.

1. UVOD

Osnovni cilj kod kodovanja govornog signala je da se signal predstavi sa minimalnim brojem bitova uz zadržavanje karakteristika da bi se dobila kvalitetna percepcija. Predložene su različite metode kodovanja kao što su LPC, CELP, RELP, VSELP, PCM, DPCM, ADPCM, Subband Coding, Transform Coding. Pobjrane metode se koriste za kompresiju, klasifikaciju, prepoznavanje govor i dr. Krajem devedesetih godina XX veka u radovima [1,2,3] predložen je novi metod za reprezentovanje signala posredstvom specijalnih funkcija koje su nazvane potpisnim funkcijama (**engl.** *Signature Function*, SF).

Osnovna ideja predloženog metoda za reprezentovanje signala zasnovana je na formiranju SF direktno iz izvora signala. U slučaju analize govornog signala vokalni trakt čoveka se smatra jednom vrstom izvora, odnosno, može se reći da svi govorni signali konstituišu specifičnu familiju signala koja se naziva 'Familija Govornih Signala' (**engl.** *Family of the Speech Signals*) [4]. U radovima [1,2,3] SF su formirane eksperimentalno. U radu [4] SF su formirane korišćenjem statističkih metoda. U cilju statističke analize govornog signala izvršen je veliki broj eksperimenata (nekoliko hiljada). U svakom eksperimentu svaki signal se deli na blokove (**engl.** *Frame*) male dužine. Za svaki frejm formirana je korelaciona matrica i određene njene sopstvene vrednosti i sopstveni vektori. Svakom frejmu pridružen je sopstveni vektor koji se bira na osnovu najveće sopstvene vrednosti. Zatim se od velikog broja sopstvenih vektora, koji je jednak broju frejmova odgovarajućeg originalnog signala, upoređivanjem sa originalnim signalom uz korišćenje norme vektora, pronalaze sopstveni vektori koji zadovoljavajuće aproksimiraju originalni signal.

Kvalitet rekonstruisanog signala ocenjuje se posredstvom MOS (**engl.** *Means Opinion Score*) testa. MOS test se sprovodi nad većim brojem slušalaca koji daju subjektivnu ocenu o kvalitetu signala.

U daljem tekstu prikazan je algoritam za reprezentovanje signala i prikazani rezultati MOS testa koji je sproveden na engleskom jeziku. U drugom delu prikazan je rezultat MOS testa koji sproveden na srpskom jeziku gde su slušaoci studenti i profesori Više tehničke škole u Nišu.

2. STATISTIČKI METOD ZA GENERISANJE POTPISNIH FUNKCIJA

Kontinualni govorni signal $x(t)$ procesom smplovanja prevodi se u diskretan signal $x(n)$. Za potrebe analize signala koristi se N uzoraka, pri čemu je N dovoljno veliki broj. Signal $x(n)$ može da se predstavi u obliku:

$$x(n) = \sum_{i=1}^N x_i \delta(n-i), \quad (1)$$

gde je $\delta(n)$ jedinični impuls, a x_i vrednost i -tog uzorka. Signal $x(n)$ deli se na frejmove dužine L_F . Nakon toga signal $x(n)$ može da se tretira kao kolekcija frejmova:

$$X_F = [X_{F1} \ X_{F2} \ \dots \ X_{FN_{FT}}], \quad (2)$$

gde je $N_{FT} = N/L_F$. Svaki frejm može se posredstvom matrice notacije tretirati kao vektor sa L_F koordinata:

$$X_{Fk} = \begin{bmatrix} x_{(k-1)L_F+1} \\ x_{(k-1)L_F+2} \\ \dots \\ x_{kL_F} \end{bmatrix}. \quad (3)$$

Dalja analiza signala realizuje se nad svakim frejmom pojedinačno. Za frejm X_{Fk} formira se matrica:

$$R_{Fk} = \begin{bmatrix} r_k(1) & r_k(2) & \dots & r_k(L_F) \\ r_k(2) & r_k(1) & \dots & r_k(L_F-1) \\ \dots & \dots & \dots & \dots \\ r_k(L_F) & r_k(L_F-1) & \dots & r_k(1) \end{bmatrix}, \quad (4)$$

gde se elementi matrice računaju prema:

$$r_k(l-1) = \frac{1}{L_F-l} \sum_{j=[(k-1)L_F+1]}^{[kL_F-1-l]} x_{j+1} x_{j+1+l}, \quad l = 0, 1, \dots, (L_F-1). \quad (5)$$

Korelaciona matrica R_{Fk} naziva se Toeplitz-ova. Sopstvene vrednosti λ_{ik} i sopstveni vektori V_{ik} računaju se iz uslova:

$$R_{Fk} V_{ik} = \lambda_{ik} V_{ik}, \quad i = 1, 2, \dots, L_F, \quad (6)$$

Koristeći matricnu notaciju sopstvene vrednosti mogu da se predstavje sa:

$$\lambda_k = [\lambda_{k1} \ \lambda_{k1} \ \dots \ \lambda_{kL_F}], \quad (7)$$

dok se sopstveni vektori mogu da predstavje kao:

$$\begin{aligned} V_{1k} &= [V_{11k} \ V_{12k} \ V_{1L_Fk}] \\ V_{2k} &= [V_{21k} \ V_{22k} \ V_{2L_Fk}] \\ &\dots \\ V_{L_Fk} &= [V_{L_F1k} \ V_{L_F2k} \ V_{L_FL_Fk}] \end{aligned} \quad (8)$$

Svakoj sopstvenoj vrednosti λ_{ki} pridružen je sopstveni vektor V_{ki} . Veza između frejma X_{Fk} i sopstvenih vektora V_{ik} može da se odredi iz:

$$X_{Fk} = \sum_{i=1}^{L_k} C_i V_{ik} \quad (9)$$

Uz uslov ortogonalnosti sopstvenih vektora:

$$V_{ik}^T V_{jk} = \begin{cases} 1 & \text{za } i = j \\ 0 & \text{za } i \neq j \end{cases} \quad (10)$$

koeficijenti C_i mogu se odrediti iz relacije:

$$C_i = X_{Fk}^T \cdot V_{ik} \quad (11)$$

Za dovoljno dug vremenski interval, odnosno, za dovoljno veliki broj frejmova N_{FT} , moguće je odrediti dužinu frejma L_F tako da se izvrši aproksimacija:

$$X_{Fk} = C_i \cdot V_{ik} \quad (12)$$

Ova aproksimacija bazira se na činjenici da je izvršeno preuređenje komponentata sopstvenog vektora u skladu sa vrednostima pridruženih sopstvenih vrednosti, tako da se formira niz vektora čija je energija u opadajućem redosledu. Saglasno tome, vektor V_{ik} je sa najvećom energijom. Uz uvažavanje iznesene pretpostavke u radu [4] je prvi sopstveni vektor nazvan *potpisnim vektorom* (**engl.** *Signature Vector*, SV). Analizom svakog frejma originalnog signala $x(n)$ moguće je formirati sekvencu potpisnih vektora S_F . Ovako potpisana sekvencu nazvana je *potpisnom sekvencom* (**engl.** *Signature Sequence*, SS).

Statističkom analizom potpisne sekvence pronalaze se određene forme, odnosno oblici (**engl.** *Shape*) prema definisanom kriterijumu. Iz mnoštva uzoraka izdvajaju se karakteristični koji su dovoljno različiti od drugih. Kao kriterijum različitosti moguće je, pored ostalog, koristiti normu vektora.

Opisanim postupkom izdvaja se M različitih potpisnih vektora iz potpisne sekvence S_F čime se formira nova potpisna sekvencu S koja se naziva *potpisnom funkcijom*. Svakom frejmu X_{FK} pridružuje se skup $\{C_{kr}, r=1, \dots, N_{FT}; S_i, i=1, \dots, M\}$. Uz činjenicu da je sekvencu S definisana, za opis frejma X_{FK} moguće je koristiti skup $\{C_{kr}, r=1, \dots, N_{FT}; i, i=1, \dots, M\}$.

3. ANALIZA EFIKASNOSTI ALGORITMA

U radu [4] izvršeno je testiranje algoritma u cilju određivanja dužine frejma L_F i dužine potpisne sekvence M . Kao mera kvaliteta rekonstruisanog signala na osnovu potpisnih funkcija korišćen je test kvaliteta čujnosti. Test je sproveden na sledeći način. Dvadeset slušalaca (10 muškog i 10 ženskog pola) slušalo je 20 rečenica. Za svaku rečenicu

svaki slušalac je na osnovu individualnog doživljenog kvaliteta dao mišljenje o kvalitetu ocenom 1-5 (**engl.** *Opinion Score*, OS). Izračunavanjem srednje vrednosti OS-a (**engl.** *Mean Opinion Score*, MOS) slušaoci su sugerisali, odnosno, klasifikovali rekonstruisani govorni signal. Klasifikacija je izvršena prema:

IF	MOS ≤ 2.2	THEN	‘loš kvalitet’
IF	2.2 < MOS ≤ 2.6	THEN	‘govor razumljiv’
IF	2.6 < MOS ≤ 3.2	THEN	‘dobar kvalitet’
IF	3.2 < MOS ≤ 4.0	THEN	‘vrlo dobar kvalitet’
IF	4 < MOS	THEN	‘izuzetan kvalitet’

Sprovedenom analizom obuhvaćeni su govorni signali do najviše:

a) 974 frejma sa dužinom frejma $L_F=24$ i dužine potpisne sekvence $M=15$ (MOS=2.762),

b) 585 frejma pri dužini frejma od $L_F=40$ i dužine potpisne sekvence $M=16$ (MOS=2.493).

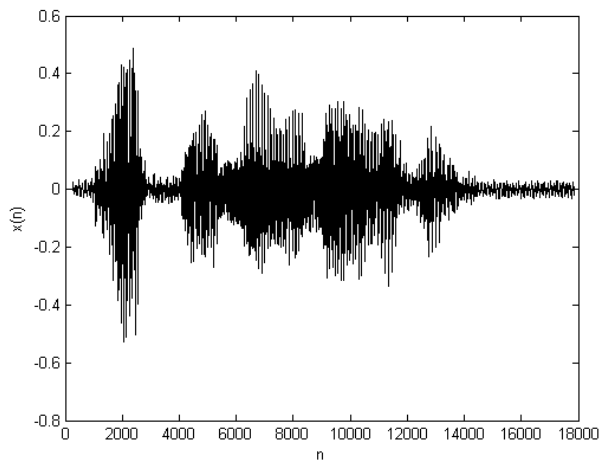
Upoređivanjem vrednosti MOS vidi se da je povoljnija situacija sa dužinom frejma $L_F=24$ i dužine potpisne sekvence $M=15$ pri kojoj je dobijen *vrlo dobar kvalitet*.

Ako se svaki frejm X_{FK} govornog signala dužine 24 odmeraka odmerava sa 8 bita po odmerku tada je ukupan broj bitova $N_b=24 \cdot 8=192$. Ovaj broj bitova može biti prenešen dalje putem telekomunikacione linije ili zapamćen. Primenom opisanog algoritma za svaki frejm potrebno je kodovati koeficijent C_{kr} pomoću 8 bita i indeks r koji ukazuje na odgovarajući potpisni vektor ($0 < r < 16$) za koga je potrebno 4 bita. Ukupan broj bitova u ovom slučaju je $8+4=12$. U ovom slučaju je kompresija (**engl.** *Compression Rate*) $CR=192/12=16$. Za slučaj kada je $L_F=40$ i dužine potpisne sekvence $M=16$ izračunava se $CR=(40 \cdot 8)/(8+4)=26.667$ uz smanjeni iznos MOS-a.

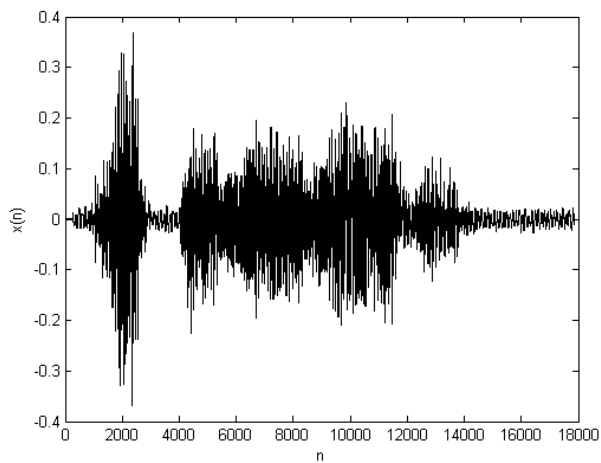
4. PRIMER PRIMENE ALGORITMA

Opisani algoritam primenjen je za kompresiju govornog signala na srpskom jeziku i nakon toga je sproveden MOS test u cilju određivanja kvaliteta reprodukovanog govornog signala. Kao izvor signala korišćeni su WAV-fajlovi u kojima su bile snimljene neke govorne sekvence iz, sa naših prostora, popularnih filmova. Za razliku od testa sprovedenog u [4] gde je bilo ograničenje broja frejmova na 974 u ovom testu sekvence govornog signala bile su promenljive dužine i kretale su se od 744-6156 frejma. Pored toga, analizirane sekvence nisu ni na koji način selektovane da bi bile pogodnije za kompresiju (samo ton spikera) već su obrađivane u originalu (buka u pozadini, više govornika, različiti zvučni efekti i dr.).

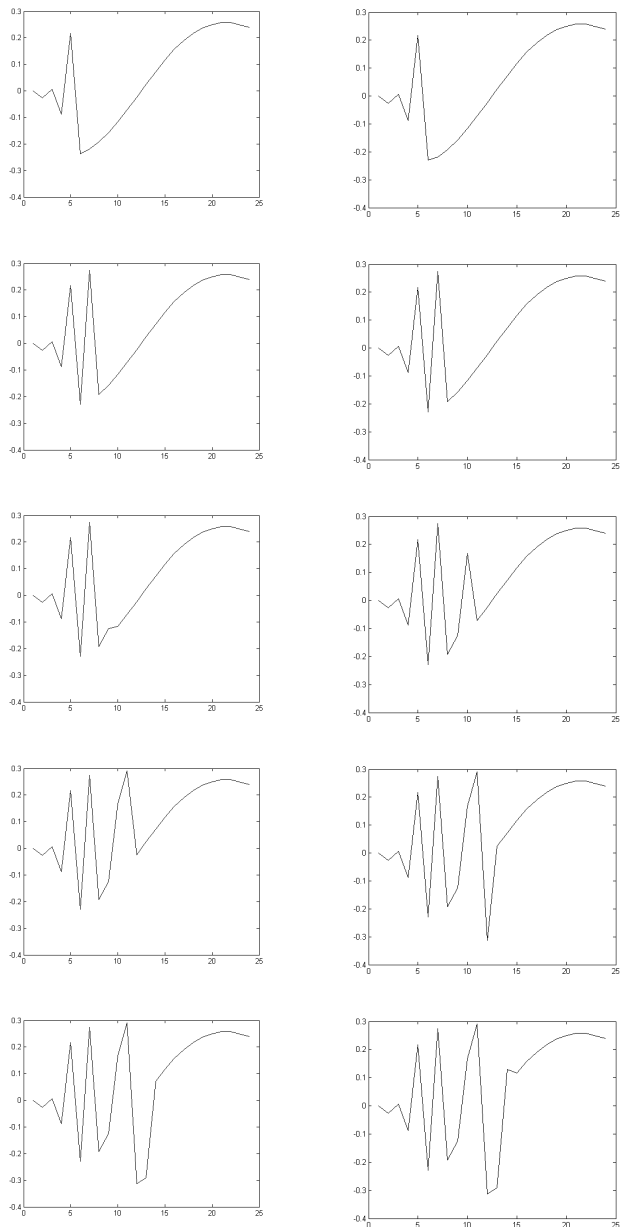
U testu su korišćeni frejmovi dužine $L_F=24$ i dužine potpisne sekvence $M=15$. Na sl.1. prikazan je vremenski dijagram originalnog signala (jednog od 20 analiziranih, rečenica ‘*Laki je malo nervozan*’) dok je na sl.2 prikazan rekonstruisani signal nakon primene opisanog algoritma za kompresiju. Na sl.3. prikazane su potpisne sekvence za isti signal.



SI.1. Originalni signal.



SI.2. Rekonstruisani signal.



SI.3. Selektovane potpisne sekvence za $LF=24$,
 $S = \{S_1, S_2, \dots, S_{15}\}$.

Za realizaciju MOS testa kao ocenjivači kvaliteta govornog signala angažovani su studenti i nastavnici Više tehničke škole u Nišu. Dvadeset slušalaca (10 muškog i 10 ženskog pola) slušalo je 20 rečenica. Za svaku rečenicu svaki slušalac je na osnovu individualno doživljenog kvaliteta dao mišljenje o kvalitetu ocenom 1-5. Izračunati koeficijent kvaliteta govornog signala iznosio je $MOS=2.4$ što je na osnovu napred opisanog kriterijuma tretirano kao *govor razumljiv*. Uzimajući u obzir samo sekvence koje su imale manje od 1000 frejma (što je slučaj sa testom sprovedenim u [4]) dobija se $MOS=2.73$ (*dobar kvalitet* govornog signala) dok je u [4] određeno $MOS=2.762$.

5. ZAKLJUČAK

U radu je prikazan rezultat MOS testa kvaliteta govornih signala na srpskom jeziku. Pre sprovođenja testa nad govornim signalom je sproveden proces kompresije prema novopredloženom algoritmu [4] koji se bazira na izračunavanju statističkih potpisnih funkcija. Rezultati testa (MOS=2.73) ukazuju na zadovoljavajući nivo kvaliteta rekonstruisanog signala.

LITERATURA

- [1] R. Akdeniz, A.M. Karas, B.S. Yarman, *Turkish speech codind by signature base sequences*, Proceedings of the ICSPAT - International Conference on Signal Processing Application & Technology, pp. 1291-1294, Toronto, Canada, Sept. 1998.
- [2] R. Akdeniz, B.S. Yarman, *Generation of optimum signature base sequences for speech signals*, Proceedings ICASSP2000 - International Conference an Acoustic, Speech and Signal Processing, Proceedings CD No: 3945, Istanbul, Turkey, June 2000.

- [3] A.M. Karas, B.S. Yarman, *A new approach for representing discrete signal waveforms via private signature base sequences*, ECCTD'95, Proceedings of Istanbul, pp. 875-878, Turkey, 1995.

- [4] R. Akdeniz, S. Yarman, *A novel method to represent speech signals*, Signal Processing, Vol. 85, pp. 37-50, 2005.

Abstract – The first part of this paper deals with an algorithm for representing of the speech signal based on signature functions. The signature functions are generated on the base of statistic parameters of the speech signal. The effect of this algorithm is analysed by means of compression factors. The second part of this paper shows the application results of the described algorithm in representing of the speech signal of the Serbian language. The results are obtained by means of the MOS test.

REPRESENTING OF THE SPEECH SIGNAL BY AN ALGORITHM WITH SIGNATURE FUNCTIONS

Z. Milivojević, M. Mirković, P. Rajković, B. Milošević