

# MULTIREZOLUCIJSKA DEKOMPOZICIJA SIGNALA GOVORA

Traljia I, Peruniia B.  
Elektrotehniiki fakultet Univerziteta u Sarajevu

## I UVOD

Do sada je uraena znaajna analiza govornih signala kao i algoritmi za njihovu detekciju, obradu i kompresiju. Tehnike i algoritmi kodiranja govornih signala uinili su brži i jednostavniji prenos signala. Algoritmi prepoznavanja govora dovele su do izdavanje komandi raunarima glasom, dok su algoritmi sinteze govora uinile jednostavniju interakciju između raunara i oovjeka.

Wavelet analiza, kao jedan od generalizovanih metoda multirezolucijske analize, relativno je nova metoda u analizi signala [1]. Za veoma kratko vrijeme našla je veoma široku primjenu. Waveleti se danas primjenjuju u digitalnim komunikacijama, procesiranju bimoedicinskih signala, procesiranju slika u medicini, astronomiji, seizmografiji, i numeriekoj analizi. Waveleti se koriste za kompresiju digitalnih signala i slika, ubrzanje osnovnih nauenih algoritama. Mnogi problemi detekcije i procjene signala sa prisustvom šuma riješeni su korištenjem waveleta [2] [3]. Waveleti sve više primjene imaju i u svakodnevnim aplikacijama na raunarima i Internetu. Najnoviji algoritam kompresije slike JPEG2000 zasnovan je upravo na waveletima .

Pored znaajne primjene za analizu raznih signala waveleti još uvijek nisu dovoljno zastupljeni u analizi govornih signala. Primjena u analizi govora može se ogledati u multirezolucijskoj analizi ispravnih i neispravnih glasova, kao i implementaciji algoritama za kompresiju govora u telekomunikacijama. Jedna od osnovnih prednosti multirezolucijske dekompozicije je analiza lokalizovane oblasti dugog signala. Takva analiza dozvoljava korištenje dugog vremenskog intervala gdje nam je potrebna preciznija informacija o niskim frekvencijama, kao i kraih intervala gdje nam je bitna informacija o višim frekvencijama. Osnovna prednost se pokazuju u analizi nestacionarnih signala, signala sa tranzijentnim ponašanjem i diskontinuitetima. Kod pogrešnog izgovora bitni su detalji, a multirezolucijska dekompozicija pokazuje odliene rezultate upravo pri analizi takvih signala. Korištenjem svih prednosti waveleta u analizi govornih signala oekivani rezultati bi si mogli primjeniti u telekomunikaciji za kompresiju glasa, u medicini za terapiju govora i slieno.

U radu je data analiza samoglasnika korištenjem multirezolucijske dekompozicije. Uzeti uzorci govornih signala su prevedeni na raunar u digitalnu formu. Zatim su pomoau MATLAB Wavelet Toolbox-a određivani wavelet koeficijanata [4]. Pri tome se vodi rauna o izboru najboljeg seta wavelta kao i odabiru koeficijenata za odabrani tip

waveleta. Izvršena je analiza nepravilno izgovorenih glasova i određeni njihovi koeficijenti. Poređenjem ovih koeficijenata pokušava se naai zavisnost kao i mogua korekcija.

U narednom poglavlju date su neke osnovne karakteristike govornih signala. U treem poglavlju, u cilju boljeg razumjevanja odabiranja koeficijenata, date su teoretske osnove wavelet analize. U etvrtom poglavlju su dati rezultati multirezolucijske analize sa skupom raspoloživih waveleta kao i izbor najboljih koeficijenata. Napravljena je analiza za svaki od samoglasnika do petog nivoa. Za svaki od samoglasnika je određen wavelet koji ima najveai varijansu. Taj wavelet za određeni koeficijent je najpogodniji za njegovu analizu.

## II KARAKTERISTIKE SIGNALA GOVORA

Govor je nestacionarni sluajni proces. Nestacionaran signal ima osobinu da su njegove statistike mijenjaju u vremenu. To je zbog toga što se i elementi koji kreiraju glas tj. oblik vokalnog trakta, pozicija jezika i oblik usta mijenjaju u vremenu. Nestacionarne procese je teško modelirati.

Međutim, ljudski govor se može posmatrati kao stacionaran u kratkom vremenskom periodu. Taj period iznosi od 20 do 40 milisekundi. Ova pretpostavka je ispravna jer je sistem za proizvodnju govora mehaniiki sistem koji se sporo mijenja u vremenu. Karakteristike signala se mogu smatrati konstantnim duž ovog vremenskog perioda.

Zbog svega navedenog govor se uobičajeno analizira u segmentima dužine od 20 do 40 ms koji se nazivaju okviri. U svakom okviru glas se tretira kao stacionaran.

Vazdušna struja se potiskuje iz pluaa. Vazdušna struja potom prolazi kroz završni dio traheje i larinks, a potom kroz ždreonu i usnu ili nosne šupljine. Ta struja je ustvari glavni zvueni talas. Modulacijom ovog talasa se formira akustieno polje koja odgovaraju osobinama izgovorenih glasova. Postoje dva razliita ponašanja ove vazdušne struje.

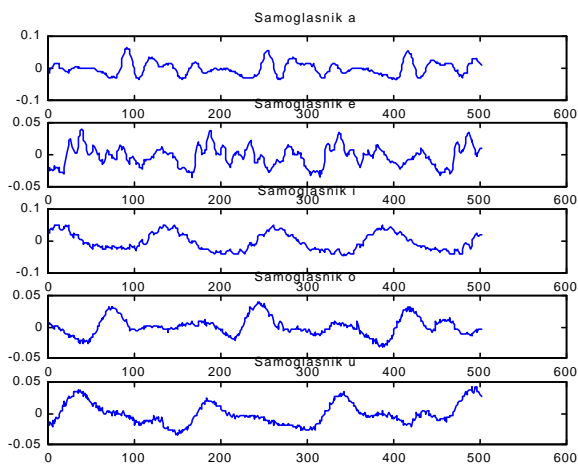
U prvom slučaju vazdušna struja može prolaziti slobodno kroz ždreonu i usnu ili nosne šupljine. Pri tome ona ne nailazi na prepreke koje bi izazvale trenje vazduha odnosno pojavu bilo kakvog šuma. Ovo je karakteristika samoglasnika. Zbog odsustva frikcionog elementa kod samoglasnika akustiika energija se ponaša periodieo.

U drugom slučaju vazdušna struja nailazi na razne prepreke koje prouzrokuju trenje vazdušnih eestica. U tom slučaju dobivamo izarazito neperiodieke oblike akustiike energije.

Sada komponente šuma, modulirane na razne načine formiraju glasove. Ovo je karakteristično za suglasnike.

Vokali se u strukturi jezika javljaju sa 41.89%. To pokazuje koja je važnost upravo u obradi upravo ovih pet glasova, što je slučaj i u ovom radu.

Od svih samoglasnika najčešće se pojavljuje samoglasnik a, zatim i, e o i najmanje u. Na slici 1 su dati dijagrami samoglasnika. Za sve njih se vidi skoro periodička komponenta.



Slika 1: Samoglasnici posmatrani na jednom periodu

### III WAVELET ANALIZA

Furijerova transformacija svoje nedostatke pokazuje u slučaju nestacionarnih procesa sa diskontinuitetima. Neki od nedostataka su prevaziđeni sa tzv. Furijerovom transformacijom u kratkom vremenu (Short Time Furier Transformation). Od nedavno, pojavio se novi koncept u analizi signala, koji svoje najbolje osobine pokazuje pri analizi nestacionarnih procesa sa diskontinuitetima[5]. Riječ je o wavelet transformaciji.

Wavelet transformaciju možemo posmatrati kao dekompoziciju signala u skup baznih funkcija (waveleta), analogno korištenju sinusa i kosinusa u Furijerovoj transformaciji za predstavljanje drugih funkcija. Ove bazne funkcije su dobivene skaliranjem i translacijom tzv. majke waveleta. Glavna razlika između wavelet transformacije i Furijerove analize je mogućnost lokalizacije vremena i frekvencije.

Signal izražen preko wavelet ekspanzije ima oblik:

$$f(t) = \sum_k \sum_j a_{j,k} \mathbf{y}_{j,k}(t)$$

Skup ekspanzionih koeficijenata  $a_{j,k}$  se naziva diskretna wavelet transformacija (Discrete Wavelet Transformation DWT) a ekspanzija  $f(t)$  predstavlja inverznu transformaciju. Wavelet ekspanzioni skup nije jedinstven. Postoje mnogi različiti skupovi waveleta. Za sve te skupove je

karakteristično da predstavljaju vremensko-frekventnu lokalizaciju signala. Broj operacija za generalni oblik wavelet transformacije je  $O(N \log(N))$ , što je isti broj kao i za brzu Furijerovu transformaciju.

Prva generacija u wavelet sistemu je generisana iz jedne funkcije skaliranja ili waveleta jednostavnim skaliranjem i translacijom. Iz takozvane majke waveleta dobiva se dvodimenzionalna parametrizacija

$$\mathbf{y}_{j,k}(t) = 2^{j/2} \mathbf{y}(2^j t - k) \quad j, k \in \mathbb{Z}$$

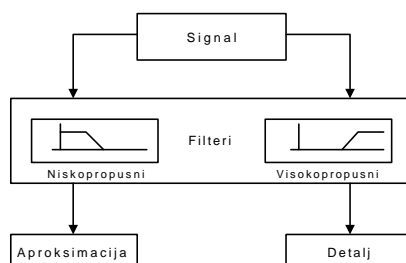
Multirezolucijska formulacija zahtjeva definisanje još jedne funkcije, koja se naziva funkcija skaliranja  $\phi(t)$ . Koristeći kombinaciju funkcije skaliranja i waveleta signal se može predstaviti i kao

$$f(t) = \sum_{k=-\infty}^{\infty} c_j \mathbf{j}(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \mathbf{y}(2^j t - k)$$

Mayer i Mallat su razvili teoriju za konstruisanje baze ortonormalnih waveleta kroz multirezolucijsku aproksimaciju signala. Taj rad je bio inspirisan radom u kompjuterskoj viziji za analizu slike na različitim rezolucijama. Dataljnija analiza osobina ortogonalnih waveleta i multirezolucijske analize dovela je do iznenađujuće veze između teorije kvadraturnih filtera u konstrukciji tzv. filter banks.

Svaki signal u sebi nosi određene frekventne sadržaje. Za niskofrekventni sadržaj signala možemo reći da on daje identitet samom signalu, dok visokofrekventna komponenta daje nijanse. Ako uzmemo ljudski govor i uklonimo visoke frekvencije, glas æ zvučati drukčije, ali æ i dalje biti razumljiv. Ako uklonimo niskofrekventnu komponentu i ostavimo samo visokofrekventni sadržaj, dobit æmo nerazumljiv govor.

Tako u wavelet analizi, odnosno u razlaganju signala govorimo o aproksimaciji i detalju. Aproksimacija je niskofrekventna komponenta a detalj je visokofrekventna komponenta. Najosnovniji princip je dat na slici 2.



Slika 2: Princip multirezolucijske analize

Osnovni cilj pri multirezoluciji je odabir dobrih koeficijenata filtera odnosno odabir dobrih waveleta.

Diskretna wavelet transformacija razmatra dva skupa funkcija. Prvi je funkcija skaliranja koja odgovara niskopropusnom filteru  $h[n]$ , a druga wavelet funkcija koja odgovara visokopropusnom filteru  $g[n]$ . Koeficijenti skaliranja, odnosno niskopropusnog filtra se mogu dobiti iz funkcije skaliranja kao:

$$h(n) = \sqrt{2} \int_{-\infty}^{\infty} \mathbf{j}(t) \mathbf{j}(2t - n) dt$$

Koeficijenti waveleta, odnosno visokopropusnog filtra se dobili iz funkcije skaliranja kao:

$$g(n) = \sqrt{2} \int_{-\infty}^{\infty} \mathbf{y}(t) \mathbf{j}(2t - n) dt$$

Koeficijenti  $h(n)$  moraju zadovoljavati slijedeće uslove

$$\sum_n h(n) = \sqrt{2}$$

$$\sum_n |h(n)|^2 = 1$$

$$\sum_n h(2n) = \sum_n h(2n + 1) = \frac{1}{\sqrt{2}}$$

$$\sum_n h(n)h(n - 2k) = \mathbf{d}(k)$$

$$h(n) = 0 \quad \text{za} \quad n < 0 \quad \text{i} \quad n > N - 1$$

Filteri nisu nezavisni, već su međusobno vezani relacijom (Quadrature Mirror Filters):

$$g[N - 1 - n] = (-1)^n h[n]$$

gdje je  $N$  dužina filtera.

Ova dekompozicija polovi vremensku rezoluciju jer samo pola uzoraka sada karakteriše polazni signal. Ali istovremeno ova operacija udvostručava frekventnu rezoluciju, jer signal sada pokriva samo pola frekventnog opsega polaznog signala. Na slici 2 je prikazan samo prvi nivo dekompozicije. Dekompozicija se nastavlja dalje do željenog nivoa. Naime, signal aproksimacije se dovodi na sljedeći par niskopropusnog i visokopropusnog filtra. Poslije svakog filtriranja polovina uzoraka može biti odbađena, jer prema Nyquistovom kriteriju signal tada ima graničnu frekvenciju  $\pi/2$  radijana, a ne  $\pi$ . Zbog toga signal može biti decimiran poslije svakog filtra, tako da se svaki drugi uzorak izostavi. Matematički izraženo to izgleda ovako:

$$y_h[k] = \sum_n x[n] \cdot g[2k - n]$$

$$y_l[k] = \sum_n x[n] \cdot h[2k - n]$$

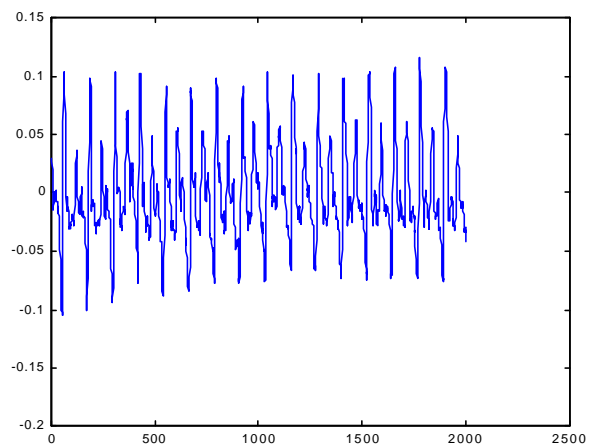
U datim jednačinama  $y_l[k]$  i  $y_h[k]$  predstavljaju izlaz iz niskopropusnog i visokopropusnog filtera, respektivno, poslije decimiranja sa dva. Ovaj proces je poznat kao "subband coding".

#### IV ODABIR KOEFICIJENATA

U ovom poglavlju dat je prikaz načina odabira wavelet koeficijenata. Vršena je analiza samo na uzorcima samoglasnika. Uzorci su uzeti od osoba sa nemogućnošću pravilnog izgovora. Ukupno je uzeto 17 uzoraka. Također je uzeto 5 uzoraka od osoba koje pravilno izgovaraju samoglasnike. Za snimljene samoglasnike su zatim određeni wavelet koeficijenti koristeći raspoložive wavelete u Matlab Wavelet Toolbox paketu. Da bi se optimizirao broj koeficijenata koji se razmatra u kvalifikaciji pojedinih glasova izvršen je također i izbor koeficijenta aproksimacije i detalja koji nose najviše informacije. Izbor najboljih waveleta je napravljen u skladu sa vrijednošću varijanse za pojedine wavelete. Koeficijenti sa maksimalnom varijansom imaju najviše rasipanje tj. najveću razliku između pojedinih uzoraka su upravo kod tih koeficijenata. Zato je i najbolje ove koeficijente uzimati u algoritmima, bilo da se radi o kompresiji glasa u komunikacijama, prepoznavanju glasa pomoću waveleta ili u terapiji govora. Glasovi su snimani prilikom logopedске terapije. Izvršeno je snimanje uzoraka na PC računaru. Na slici 3 dat je prikaz jednog snimljenog samoglasnika.

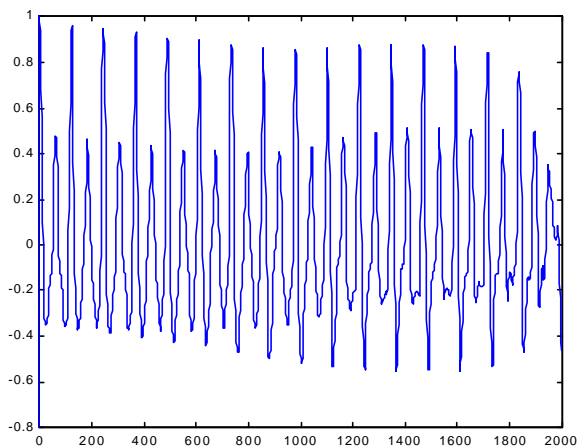
Za analizu digitalnog uzorka govornog signala, kao i za wavelet analizu neophodno je bilo koristiti dovoljno snažan i efikasan software. U ovom radu je korišten program MATLAB sa Wavelet Toolbox i Neural Networks Toolbox.

Očigledna je skoro periodička komponenta posmatranog uzorka samoglasnika 'A' datog na slici 3. Da bi se odredio period nađena je autokorelaciona funkcija. Razlika između maksimuma te autokorelacione funkcije predstavlja period.



Slika 3: Skoro periodička komponenta samoglasnika

Autokorelacina funkcija za dati signal ima izgled kao na slici 4. Napravljena je funkcija koja određuje period između maksimuma autokorelacione funkcije



Slika 4: Autokorelacina funkcija za ispitani uzorak

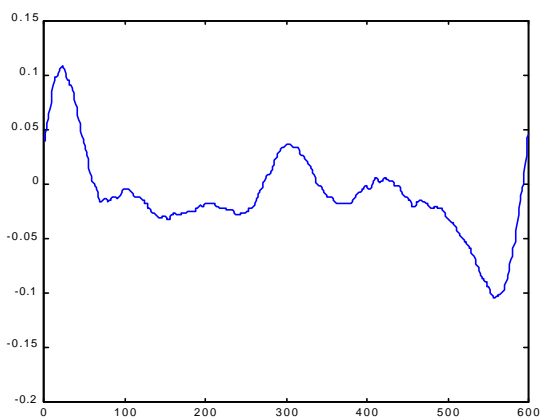
Za dati primjer dobije se da je period 132, što za datu frekvenciju uzimanja uzoraka daje dužinu 299 ms. Za daljnju analizu uzima se samo signal na tom osnovnom periodu. Za svaki uzorak je kao početak uzet prolazak kroz nulu prije maksimuma uzorka.

Na navedenim uzorcima za svaki samoglasnik i svaki uzorak izvršene su sljedeće modifikacije:

-izvršena je interpolacija signala na period 600, tako da se svi signali posmatraju na istom periodu

-vrijednosti uzoraka su preračunate tako da svi posmatrani uzorci imaju istu energiju; na taj način jačina izgovorenog signala ne utiče na vrijednosti koeficijenata u multirezolucijskoj analizi i pronalaženju koeficijenata waveleta.

Dati uzorak, pošto je sveden na period 600 i vrijednosti uzoraka izračunate tako da je energija uzorka 1 ima izgled kao na slici 5.



Slika 5: Signal na osnovnom periodu

Za db5 wavelete ukupan broj uzoraka signala je 600, a broj koeficijenata za aproksimaciju i svaki od 5 detalja dat je u tabeli I.

Nivo	Broj koeficijenata
Aproksimacija A5	27
Detalj D5	27
Detalj D4	45
Detalj D3	82
Detalj D2	156
Detalj D1	304

Tabela I: Broj koeficijenata po koracima analize

S obzirom da se ovde dekompozicija radila do petog nivoa aproksimacija A5 su koeficijenti  $a_{j,k}$   $j=5, k=1, \dots, 29$ , a koeficijenti detalja su  $d_{j,k}$   $j=1, 2, \dots, 5$ . Koeficijenti predstavljaju izlaze iz odgovarajućih filtera. Pošto je ulazni govorni signal uzorkovan sa frekvencijom 22050 Hz, granična frekvencija signala prije wavelet analize je 11025 Hz. Odgovarajući frekventni opsezi za koeficijente dati su u tabeli II. Frekventni opsezi su isti i za bior5.5 i db5 wavelete, ali se radi o različitim filterima, tj. o koeficijentima odgovarajućih nisko i visokopropusnih filtera.

Koeficijenti	Frekventni opseg (Hz)
$d_{1,k}$	5512-11025
$d_{2,k}$	2756-5512
$d_{3,k}$	1378-2756
$d_{4,k}$	689-1378
$d_{5,k}$	344-689
$a_{5,k}$	0-344

Tabela II: Veza između koeficijenata i opsega filtra. Sljedeći korak je bio da se napravi multirezolucijska analiza svih samoglasnika sa svim raspoloživim waveletima koji su ponuđeni. MATLAB Wavelet Toolbox sadrži sljedeće wavelete za višenivovsku 1-D wavelet analizu: Haarovi waveleti (isti kao 'db1' waveleti), devet preostalih iz familije Daubechies waveleta, 14 biortogonalnih waveleta, pet Coiflet i sedam Symlet waveleta.

Napravljena je dekompozicija korištenjem svih navedenih waveleta. U multirezolucijskoj analizi išlo se do petog nivoa detalja. To se pokazuje kao sasvim dovoljno jer se vidi da su većine koeficijenata detalja na petom nivou jednake ili veoma blizu nuli.

Potom su određene vrijednosti varijansi za svaki od koeficijenata za pojedine wavelete i izračunate srednja vrijednost. Za daljnju analizu bilo koga od glasova trebalo bi uzeti koeficijente izračunate sa najpogodnijim waveletima, odnosno onim koji imaju najveću vrijednost. Prilikom analize se pokazalo da se najbolji rezultati dobivaju sa biorovim waveletima.

Koeficijent	cA5	cD5	cD4	cD3	cD2	cD1
Ispravno	bior3.1	bior1.3	db1	bior1.5	bior1.3	db1
Neispravno	bior3.1	db1	bior3.1	bior3.5	bior1.3	db1

Tabela III: Maksimalne varijanse za pojedine koeficijente za slovo A

## V DISKUSIJA I ZAKLJUČCI

U radu je dat prikaz wavelet analize, kao i metod izbora najboljih koeficijenata za signale govora. Pri tome su u analizi uzeti glasovi koje nose skoro periodičku komponentu, tj. samoglasnici.

Metod odabira najboljih koeficijenta baziran je na najvećoj vrijednosti varijanse. Za analizirane glasove je izvršena multirezolucijska analiza koristeći raspoložive waveleta. Za glasove je određen najbolji wavelet za svaki koeficijent, tj. nivo analize. Bior waveleti se pokazuju kao najbolji za ovaj tip signala.

Rezultati mogu imati široku primjenu. Jedan od vidova primjene je u kompresiji signala govora u telekomunikacijama. Također je moguće odabrane koeficijente uvoditi u vještačku neuronsku mrežu i vršiti kvalifikaciju u skladu sa izborom mreže. To može biti prepoznavanje samoglasnika ili klasifikacija neispravnih i ispravnih glasova.

## LITERATURA

[1]C. Sidney Burrus, Ramesh A. Gopinath, Haitao Guo, Introduction to Wavelets and Wavelet Transforms, Prantice Hall, 1988

[2]Henry M. Polchlopek and Joseph P. Noonan, Wavelets, Detection, Estimation, and Sparsity, Digital Signal Processing, 7 1997

[3]Andrew Bruce and David Donoho and Hong-Ye Gao, Wavelet analysis, IEEE Spectrum, October 1996

[4]Wavelet Toolbox User's Guide, Mathworks Inc., 1997

[5]Xiaowei Xu, Speech Compression Using Wavelets, Msc. Thesis, Lamar University May, 1998

**Abstract:** The main principles of the multiresolution decomposition of the speech signals are presented. Some basic theory fact about wavelets are presented. The best wavelets for each voice sound and each level of decomposition are selected. Also, for each signal, defected and correct signals are analyzed. Results can be used in telecommunication for compression of the voice signal and for algorithms of voice recognition.

**MULTIRESOLUTION OF THE SPEECH SIGNALS**  
**Traljic I., Perunić B.**