SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 2108

MJERE UDALJENOSTI U OBRADI GOVORNOG SIGNALA

Ivor Rendulić

Zagreb, lipanj 2011.

ZAHVALA

prof. dr. sc. Davoru Petrinoviću,

na svoj pruženoj pomoći pri izradi ovog rada,

i jer me dodatno zainteresirao za područje obrade govora.

Sadržaj

Uvod

Ideja govorom upravljanih računalnih sustava, poput brodskog računala u *Zvjezdanim stazama*, odavno je prisutna i motivira znanstvenike i inženjere da istražuju područje obrade i automatskog prepoznavanja govora te pokušaju razviti sustave koji će uspješno prepoznati ljudski govor. Desetljeća istraživanja dovela su do tri različita pristupa prepoznavanju govora (kronološki):

- akustično-fonetski, temeljen na teoriji o postojanju konačnog broja fonema
- prepoznavanje uzoraka iz predložaka, temeljeno na parametrizaciji i uspoređivanju sličnosti govornih signala
- statistički, temeljen na skivenim Markovljevim modelima

U pristupu prepoznavanja uzoraka iz predložaka za objektivno se određivanje sličnosti koriste mjere udaljenosti govornog signala. Njihova je zadaća matematički prikazati sličnost dvaju govornih signala onako kako bi to učinio i slušatelj – za slične signale rezultat treba biti mala udaljenost, a za različite velika. S obzirom na složenost područja obrade govora, radi se o nimalo jednostavnom problemu. Potrebno je prvo parametrizirati govorni signal na način najpogodniji za usporedbu, a zatim i pronaći mjeru udaljenosti koja će zadovoljavati spomenuti subjektivni dojam.

Osim u svrhu prepoznavanja govora, mjere udaljenosti koriste se i za vrednovanje kvalitete kodiranja govora tzv. *vocoderima*, kod kojih nije važno da kodirani signal bude istog valnog oblika, već isključivo da kodirani signal zvuči što sličnije originalnom.

Cilj je rada opisati neke najčešće korištene mjere udaljenosti govornog signala imajući u vidu njihovo korištenje u svrhu prepoznavanja govora, te usporediti njihovu učinkovitost i računalnu složenost. Budući da gotovo sve mjere udaljenosti za parametrizaciju govornog signala koriste LPC ili kepstralne koeficijente, prvi će dio rada biti posvećen LPC i kepstralnoj analizi govora.

1. Prepoznavanje uzoraka govora iz predložaka

Do šezdesetih godina dvadesetog stoljeća u području prepoznavanja govora prevladavao je akustično-fonetski pristup prema kojem postoji konačan broj fonema. Svaki fonem je karakteriziran određenim svojstvima u spektru i za promatrani uzorak govora se izravno na temelju izmjerenih karakteristika odlučivalo o kojem se fonemu radi. Razvojem koncepta linearnog prediktivnog kodiranja kao metode parametrizacije govora, korištenjem dinamičkog programiranja za vremensko poravnavanje i određivanjem sličnosti dvaju govornih signala nekom od mjera udaljenosti govora polako je došlo do promjene u osnovnoj ideji prepoznavanja. Umjesto odlučivanja temeljem izmjerenih svojstava, govor koji je potrebno prepoznati uspoređivao se s referentnim, otprije poznatim i parametriziranim uzorcima koji se dobivaju postupkom "treniranja" – više testnih uzoraka iste klase se parametrizira i izračuna se najbolja parametrizacija za tu klasu. Postupak treniranja ponovi se za sve klase uzoraka te se izgradi baza uzoraka s kojima se ulazni govor uspoređuje. Takav pristup prepoznavanju govora naziva se prepoznavanjem uzoraka iz predložaka.

2. Linearno prediktivno kodiranje govora

Analiza govora temeljena na linearnoj predikciji (*Linear Predictive Coding* – LPC) najčešće je korištena tehnika za određivanje osnovnih značajki govornog signala (spektar, frekvencije formanata, funkcija vokalnog trakta, model i spektralna ovojnica signala, pobudni signal), kao i za kodiranje s postizanjem visoke kompresije u svrhu prijenosa ili pohrane govornog signala. Neke od prednosti korištenja LPC-a u analizi govornog signala te u svrhu prepoznavanja govora su:

- odlična aproksimacija govornog signala, posebice za zvučne glasove
- odvajanje modela vokalnog trakta od pobudnog signala te dobivanje spektralne ovojnice
- jednostavnost hardverske ili softverske implementacije

Za primjenu u mjerenju udaljenosti između govornih signala od posebne je važnosti druga točka, mogućnost dobivanja prijenosne funkcije vokalnog trakta i spektralne ovojnice signala. U poglavlju o mjerama udaljenosti bit će objašnjeno zašto je poželjno izdvojiti upravo vokalni trakt, a eliminirati pobudni signal.

2.1. LPC model

Ideja linearnog prediktivnog kodiranja je aproksimacija uzorka govornog signala $s(n)$ linearnom kombinacijom p prethodnih uzoraka

$$s(n) \approx \sum_{i=1}^{p} a_i \, s(n-i), \tag{2.1}$$

gdje je p red prediktora, a koeficijenti a_i su pretpostavljeni konstantnima unutar analiziranog područja. Uključi li se i pobudni signal $Gu(n)$, gdje je $u(n)$ normalizirani signal pobude, a G pojačanje, gornji izraz postaje jednakost

$$s(n) = \sum_{i=1}^{p} a_i \, s(n-i) + Gu(n) \, . \tag{2.2}$$

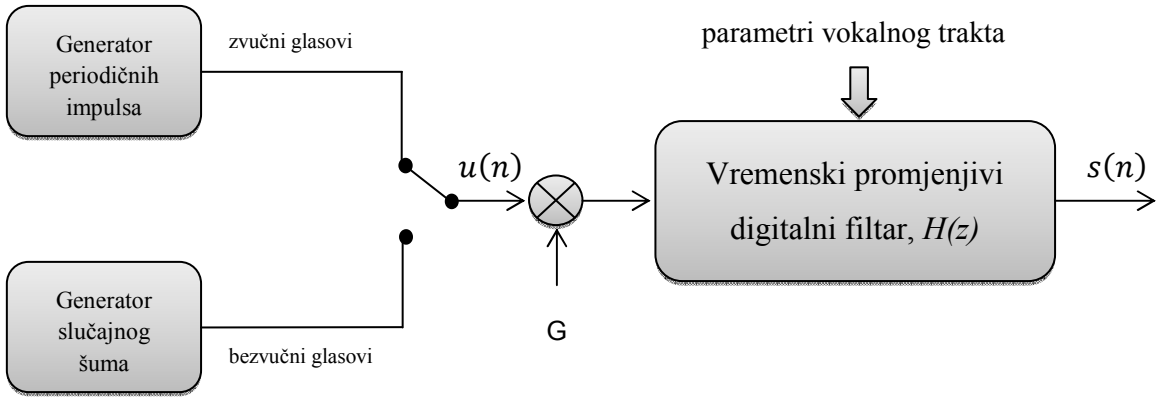
Primjenom *z*-transformacije

$$S(z) = \sum_{i=1}^{p} a_i \, z^{-i} S(z) + GU(z) \tag{2.3}$$

dolazi se do prijenosne funkcije vokalnog trakta:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{i=1}^{p} a_i \, z^{-i}} = \frac{G}{A(z)}, \tag{2.4}$$

gdje je $A(z)$ prijenosna funkcija pogreške predikcije o kojoj će biti više govora kasnije. Budući da je pobudni signal govora $u(n)$ periodički niz impusa u slučaju zvučnih glasova, odnosno slučajni šum kod bezvučnih, a prijenosna funkcija $H(z)$ opisuje vokalni trakt, model procesa sinteze govora može se prikazati blok dijagramom na slici 1.



Slika 1. Blok dijagram sinteze govora

Parametri modela su klasifikacija zvučno/bezvučno, period impulsa pobudnog signala za zvučne glasove, pojačanje normaliziranog pobudnog signala i koeficijenti digitalnog filtra a_i. Svi

parametri vremenski su promjenjivi, ali, ako promatramo kratak interval govornog signala, mogu se smatrati konstantnima.

2.2. Minimizacije pogreške predikcije

Problem analize linearnom predikcijom jest kako iz govornog signala odrediti koeficijente a_i. Osnovni pristup za računanje koeficijenata je minimizacija očekivanja kvradrata signala pogreške prediktora. Promatrat će se segmenti govornog signala koji započinju sa uzorkom n te se stoga uvode oznake oblika

$$x_n(m) = x(n+m)$$

Ako je $e_n(m)$ pogreška prediktora

$$e_n(m) = s_n(m) - \tilde{s}_n(m), \tag{2.5}$$

gdje je sa $s_n(m)$ označena vrijednost uzorka signala, a $\tilde{s}_n(m)$ je predikcija tog uzorka

$$\tilde{s}_n(m) = \sum_{i=1}^{p} a_i \, s_n(n-i), \tag{2.6}$$

cilj je minimizirati ukupnu pogrešku

$$E_n = \sum_{m} e_n^2(m) \, . \tag{2.7}$$

Uvrštavanjem izraza za (2.5) i (2.6)u gornji izraz dobije se:

$$E_n = \sum_{m} \left[s_n(m) - \sum_{k=1}^{p} a_k \, s_n(m-k) \right]^2 . \tag{2.8}$$

Za određivanje minimuma tog izraza potrebno ga je parcijalno derivirati po svim koeficijentima a_i i izjednačiti s nulom:

$$\frac{\partial E_n}{\partial a_i} = 0, \qquad 1 \leq i \leq p. \tag{2.9}$$

Parcijalnom derivacijom dobije se:

$$\frac{\partial E_n}{\partial a_i} = \sum_m 2\left[s_n(m) - \sum_{k=1}^{p} \hat{a}_k\, s_n(m-k)\right][-s_n(m-i)] = 0 \tag{2.10}$$

te uz daljnje sređivanje:

$$\sum_m s_n(m)s_n(m-i) - \sum_m \sum_{k=1}^{p} \hat{a}_k\, s_n(m-k)\, s_n(m-i) = 0, \qquad 1 \leq i \leq p. \tag{2.11}$$

Sa $\hat{a}_k$ su sada označeni optimalni LPC koeficijenti uz koje je minimizirana pogreška predikcije. Zapisano na malo drugačiji način:

$$\sum_m s_n(m)s_n(m-i) = \sum_{k=1}^{p} \hat{a}_k \sum_m s_n(m-i)s_n(m-k) = 0, \qquad 1 \leq i \leq p. \tag{2.12}$$

Uz pretpostavku ergodičnosti procesa, u gornjem se izrazu mogu prepoznati autokorelacije:

$$\phi_n(i,k) = \sum_m s_n(m-i)s_n(m-k) \tag{2.13}$$

te dobiti kraći zapis:

$$\phi_n(i,0) = \sum_{k=1}^{p} \hat{a}_k \phi_n(i,k), \qquad 1 \leq i \leq p. \tag{2.14}$$

Očito je za određivanje optimalnih koeficijenata $\hat{a}_k$ potrebno izračunati $\phi_n(i,k)$ za $1 \leq 1 \leq p$ i $00 \leq k \leq p$ te zatim riješiti p jednadžbi sa p nepoznanica (red prediktora p je obično između 8 i 14). Za rješavanje se najčešće koristi metoda autokorelacija opisana u nastavku.

2.3. Računanje LPC koeficijenata

Budući da se koeficijenti a_i moraju iznova računati za svaki segment govornog signala, važno je paziti koliko je efikasna metoda kojom ih računamo u vidu brzine i potrebne memorije za računanje. Najčešće korištene metode su kovarijancijska i učinkovitija autokorelacijska, a uz njih će ukratko biti opisana i rešetkasta metoda.

2.3.1. Kovarijancijska metoda

Kovarijancijska metoda računa koeficijente a_i izravno iz izraza (2.14). Promatra se segment govornog signala duljine N i minimizira pogreška predikcije

$$E_n = \sum_{m=0}^{N-1} e_n^2(m).$$

Funkcija $\phi_n(i,k)$ tada je (pomakom za $-i$) definirana kao

$$\phi_n(i,k) = \sum_{m=-i}^{N-i-1} s_n(m)s_n(m+i-k)$$

te je izraz (2.14) moguće je zapisati matrično

$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \cdots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \cdots & \phi_n(2,p) \\ \phi_n(3,1) & \phi_n(3,2) & \phi_n(3,3) & \cdots & \phi_n(3,p) \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_n(p,1) & \phi_n(p-2) & \phi_n(p-3) & \cdots & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \phi_n(3,0) \\ \vdots \\ \phi_n(p,0) \end{bmatrix}. \quad (2.15)$$

Budući da vrijedi $\phi_n(a,b) = \phi_n(b,a)$, matrica s lijeve strane je simetrična i za rješavanje ovog sustava matričnih jednadžbi koristi se Choleskyjeva dekompozicija opisana u [1].

2.3.2. Autokorelacijska metoda

Za određivanje optimalnih LPC koeficijenata a_i najčešće se koristi autokorelacijska metoda. Iz izraza za prijenosnu funkciju $H(z)$ vidi se da koeficijenti a_i određuju lokacije polova i da je potrebno provjeriti stabilnost sustava za svake izračunate vrijednosti koeficijenata. Jedna od

prednosti autokorelacijske metode je da daje koeficijente uz koje je $H(z)$ sigurno stabilan i tako eliminira potrebu za provjerom stabilnosti. Osim toga, pokazat će se da autokorelacijska metoda ima još neka zgodna matrična svojstva koja olakšavaju i ubrzavaju proračun.

Promatra se segment konačne duljine *N* uzoraka te se pretpostavlja da je jednak nuli izvan tog segmenta. Iz tog se razloga signal $s_n(m)$ množi sa $w(m)$, funkcijom čija je vrijednost različita od nule samo unutar promatranog segmenta

$$s_n(m) = \begin{cases} s(m+n)w(m), & 0 \le m \le N-1 \\ 0, & inače. \end{cases} \tag{2.16}$$

Ukupna pogreška tada iznosi

$$E_n = \sum_{m=0}^{N+p-1} e_n^2 \tag{2.17}$$

i traje *p* uzoraka dulje od promatranog uzorka signala. Razlog tomu je što će se uzorci nakon *N*-tog, koji su jednaki nula, predviđati iz *p* prethodnih uzoraka koji su različiti od nule te će doći do pogreške predikcije. Također se uočava problem s pogreškama predikcije uzoraka na početku prozora koji se predviđaju iz uzoraka jednakih nuli. Greške predikcije na početku i na kraju segmenta ublažavaju se korištenjem "glatkih" funkcija $w(m)$ temeljenih na kosinusnim funkcijama kao što su Hannova ili Hammingova funkcija

$$w_{Hann}(m) = 0.5\left[1 - \cos\left(\frac{2\pi m}{N-1}\right)\right]$$

$$w_{Hamming}(m) = 0.54 - 0.46\cos\left(\frac{2\pi m}{N-1}\right).$$

Kako je sada vrijednost $s_n(m)$ različita od nule samo za $0 \le m \le N-1$, izraz za $\phi_n(i,k)$ može se ograničiti na konačnu sumaciju

$$\phi_n(i,k) = \sum_{m=0}^{N-1} s_n(m-i)s_n(m-k), \qquad \begin{matrix} 1 \le i \le p \\ 0 \le k \le p \end{matrix} . \tag{2.18}$$

Drukčijim zapisom

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \tag{2.19}$$

primjećuje se da je $\phi_n(i,k)$ sada postala kratkotrajna korelacija

$$\phi_n(i,k) = R_n(i-k), \tag{2.20}$$

$$R_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \ . \tag{2.21}$$

Funkcija R_n je parna pa vrijedi

$$\phi_n(i,k) = R_n(|i-k|) \tag{2.22}$$

i sada se pojavilo p jednadžbi sa p nepoznanica koje treba riješiti kako bi se odredili LPC koeficijenti

$$\sum_{k=1}^{p} \hat{a}_k R_n(|i-k|) = R_n(i), \qquad 1 \le i \le p. \tag{2.23}$$

Gornji je izraz moguće napisati i kao matričnu jednadžbu

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \cdots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \cdots & R_n(p-3) \\ \vdots & \vdots & \vdots & & \vdots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \cdots & R_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \vdots \\ R_n(p) \end{bmatrix}. \quad (2.24)$$

Vrijednosti R_n lako je odrediti sumacijom uzoraka signala, no za rješavanje ove jednadžbe i određivanje koeficijenata $\hat{a}_i$ potrebno je izračunati inverz matrice dimenzija $p \times p$ što je računski zahtjevan posao i velik problem ako je govor potrebno obrađivati u stvarnom vremenu. Međutim, radi se o Toeplitzovoj matrici, simetričnoj matrici čija se svaka dijagonala sastoji od istih koeficijenata, što uvelike olakšava i ubrzava rješavanje matrične jednadžbe korištenjem nekih iterativnih metoda. Najčešće korištena metoda temelji se na Drubinovom algoritmu koji je opisan ispod. Radi preglednosti korelacije su zapisane bez sufiksa *n* kao $R(i)$.

Početni uvjet:

$$E^{(0)} = R(0). \quad (2.25)$$

Rekurzivni postupak za $1 \leq i \leq p$:

$$k_i = \frac{\left[R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j)\right]}{E^{(i-1)}} \quad (2.26)$$

$$a_i^{(i)} = k_i \quad (2.27)$$

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}, \qquad 1 \leq j \leq i-1 \quad (2.28)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}. \quad (2.29)$$

Oznaka $a_j^{(i)}$ označava *i*-tu iteraciju *j*-tog koeficijenta, $E^{(i)}$ *i*-tu iteraciju pogreške predikcije, k_i su koeficijenti parcijalnih korelacija, tzv. PARCOR koeficijenti. Nakon p prolaza kroz rekurziju algoritam daje LPC koeficijente za prediktor reda p.

2.3.3. Rešetkasta metoda

I kovarijancijska i autokorelacijska metoda računanja LPC koeficijenata mogu se podijeliti u dvije faze - prvo je potrebno odrediti autokorelacije ϕ_n, odnosno R_n, a zatim izračunati koeficijente matričnim operacijama (najčešće Choleskyjeva dekompozicija u kovarijancijskoj te Durbinov algoritam u autokorelacijskoj metodi). Kod rešetkaste metode nema eksplicitnog računanja autokorelacija, već se LPC koeficijenti određuju pomoću pogreške predikcije.

Za shvaćanje rešetkaste metode potrebno je prethodno proučiti Durbinov algoritam primjenjen u autokorelacijskoj metodi. Nakon *i* iteracija Durbinov algoritam daje skup od *i* koeficijenata

$$\left\{a_j^{(i)}, \qquad j = 1,2, \dots, i\right\}$$

gdje *j* označava broj koeficijenta, a (*i*) *i*-tu iteraciju koeficijenta. Koristeći te koeficijente moguće je realizirati filtar $A(z)$ koji zapravo predstavlja prijenosnu funkciju pogreške predikcije. Umjesto oznaka

$$e_n^{(i)}(m) = e(n+m)w(m)$$

$$s_n(m) = s(n+m)w(m)$$

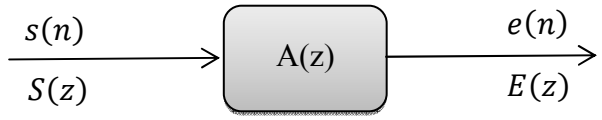
radi preglednosti koristit će se skraćene oznake $s(m)$ za govorni signal i $e^{(i)}(m)$ za *i*-tu pogrešku predikcije koja se može izraziti kao:

$$e^{(i)}(m) = s(m) - \sum_{k=1}^{i} a_k^{(i)} s(m-k). \tag{2.30}$$

U ovom je postupku važno napomenuti da je $e^{(i)}(m)$ pogreška unaprijedne predikcije jer će se kasnije pojaviti i pogreška unazadne predikcije, no zasad će se za $e^{(i)}(m)$ koristiti samo izraz pogreška predikcije. Primjenom *z*-transformacije dolazi se do izraza za prijenosnu funkciju pogreške predikcije u *i*-tom koraku

$$E^{(i)}(z) = S(z)\left[1 - \sum_{k=1}^{i} a_k^{(i)} z^{-k}\right] \tag{2.31}$$

$$A^{(i)}(z) = \frac{E^{(i)}(z)}{S(z)} = 1 - \sum_{k=1}^{i} a_k^{(i)} z^{-k}. \tag{2.32}$$



Uvrštavanjem iterativnog izraza za $a_j^{(i)}$ iz Durbinovog algoritma

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)} \tag{2.33}$$

u izraz (2.32) i nakon sređivanja se dobije:

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}). \tag{2.34}$$

Pogreška predikcije se sada može zapisati kao

$$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i z^{-i} A^{(i-1)}(z^{-1})S(z) \tag{2.35}$$

i sastoji se od dva člana. Prvi član je pogreška predikcije prethodnog koraka $E^{(i-1)}(z)$, dok se drugi član može zapisati preko novog izraza $B^{(i)}(z)$ definiranog kao

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1})S(z). \tag{2.36}$$

Primjenom inverzne *z*-transformacije na izraz (2.36) u vremenskoj se domeni dobiva:

$$b^{(i)}(m) = s(m-i) - \sum_{k=1}^{i} a_k^{(i)} s(m+k-i). \tag{2.37}$$

Postoji očita sličnost između gornjeg izraza i izraza za pogrešku predikcije $e^{(i)}(m)$. Potonji predstavlja pogrešku u predviđanju uzorka $s(m)$ iz uzoraka koji mu prethode, odnosno pogrešku unaprijedne predikcije. U slučaju $b^{(i)}(m)$ radi se o ranije spomenutoj pogrešci unazadne predikcije - uzorak $s(m-i)$ se predviđa iz uzoraka koji ga slijede. Korištenjem i unaprijedne i unazadne predikcije iz istog se niza uzoraka predviđaju i budući i prošli uzorci.

Korištenjem izraza (2.36) unaprijedna pogreška predikcije može se napisati kao

$$E^{(i)}(z) = E^{(i-1)}(z) - k_i z B^{(i-1)}(z), \tag{2.38}$$

odnosno u vremenskoj domeni

$$e^{(i)}(m) = e^{(i-1)}(m) - k_i b^{(i-1)}(m-1). \tag{2.39}$$

Time je pogreška unaprijedne predikcije u i-tom koraku izražena pomoću pogrešaka unaprijedne i unazadne predikcije iz koraka (i-1). Uvrštavanjem izraza (2.34) u (2.36)

$$B^{(i)}(z) = z^{-i} A^{(i-1)}(z^{-1}) S(z) - k_i A^{(i-1)}(z) S(z), \tag{2.40}$$

malim sređivanjem

$$B^{(i)}(z) = z^{-1} B^{(i-1)}(z) - k_i E^{(i-1)}(z) \tag{2.41}$$

i prikazom u vremenskoj domeni

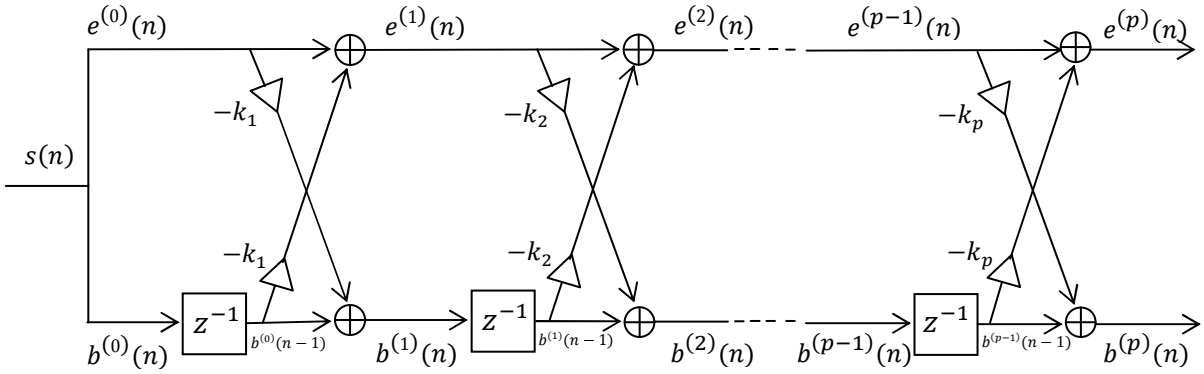
$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_i e^{(i-1)}(m) \tag{2.42}$$

vidiljivo je da se i pogreška unazadne predikcije u i-tom koraku može prikazati pomoću obje pogreške u prošlom koraku. Korištenjem izraza (2.39) i (2.42), uz poznavanje nekog početnog uvjeta, može se iterativno doći do p-tog koraka što je i konačni cilj. Budući da i zapravo

predstavlja trenutni red filtra, kao početne uvjete mogu se uzeti nulte pogreške unaprijedne i unazadne predikcije koje obje iznose $s(m)$

$$e^{(0)}(m) = b^{(0)}(m) = s(m). \tag{2.43}$$

Sada je moguće prikazati rešetkastu metodu blok dijagramom na slici 2.



Slika 2. Blok dijagram rešetkaste metode

PARCOR koeficijenti k_i računaju se iz pogrešaka predikcije u koraku (i-1) prema formuli

$$K_i = \frac{2\sum_{m=0}^{N-1} e^{(i-1)}(m)b^{(i-1)}(m-1)}{\sum_{m=0}^{N-1}[(e^{(i-1)}(m))^2+((b^{(i-1)}(m-1))^2]} \tag{2.44}$$

izvedenoj Burgovim algoritmom [5]. Uvrštavanjem (2.39) i (2.42) u (2.44) lako se pokazuje da je gornji izraz istovjetan izrazu (2.26), ali je izbjegnuto računanje autokorelacija. Iz k_i se prema (2.27) i (2.28) dobivaju traženi LPC koeficijenti a_i.

Korištenje rešetkaste metode umjesto izravne primjene Durbinovog algoritma ima prednosti u jednostavnoj, lako modularnoj blokovskoj izvedbi. Povećanjem reda prediktora p potrebno je samo dodati još jedan blok, dok ostatak sustava ostaje nepromjenjen .

2.3.4. Usporedba metoda

U tablici 1 su prikazane složenosti i potrebni podatkovni prostor za kovarijancijsku, autokorelacijsku i rešetkastu metodu. S N je označen broj uzoraka u promatranom segmentu govora, a s p red prediktora.

Tablica 1. Usporedba učinkovitosti metoda

		Kovarijancijska metoda	Autokorelacijska metoda	Rešetkasta metoda
Potrebni podatkovni prostor	Podaci	N	N	$3N$
	Matrice	$\sim p^2/2$	$\sim p$	-
	Izdvajanje promatranog segmenta	-	N	-
Računalna složenost	Računanje autokorelacija	$\sim Np$	$\sim Np$	-
	Rješavanje matričnih jednadžbi	$\sim p^3$	$\sim p^2$	$5Np$
	Izdvajanje promatranog segmenta	-	N	-

Za svaku je metodu potreban podatkovni prostor u koji se spremaju podaci – u slučaju kovarijancijske i autokorelacijske metode pohranjuje se samo N promatranih uzoraka segmenta govora, dok je za rešetkastu metodu potrebno pohraniti i unaprijednu i unazadnu predikciju, odnosno ukupno $3N$. Za razliku od rešetkaste metode koje ne računa eksplicitno autokorelacije, kod kovarijancijske i autokorelacijske metode treba pohraniti i matrice autokorelacija. Matrice imaju p^2 elemenata, ali kod kovarijancijske metode matrica je simetrična što znači da je dovoljno pohraniti $p^2/2$ elemenata. U autokorelacijskoj metodi matrica je Toeplitzova i dovoljno je pohraniti p elemenata. U autokorelacijskoj metodi provodi se izdvajanje promatranog segmenta govora množenjem sa funkcijom $w(n)$ i treba pohraniti i N njenih uzoraka.

Što se tiče računalne složenosti, kovarijancijska i autokorelacijska metoda imaju složenost proporcionalnu sa Np za računanje autokorelacije, dok se u rešetkastoj metodi autokorelacije ne računaju. U autokorelacijskoj metodi složenost izdvajanja promatranog segmenta množenjem sa

$w(n)$ je N, dok u druge dvije metode to nije potrebno. Rješavanje matrične jednadžbe Choleskyjevom dekompozicijom u kovarijancijskoj metodi rezultira složenošću proporcionalnom p^3. Autokorelacijska metoda i rješavanje Durbinovim algoritmom daju složenost proporcionalnu p^2, a za rešetkastu se metodu može pokazati da je složenost jednaka $5Np$.

Iz ove je usporedbe vidljiva očita prednost autokorelacijske nad kovarijancijskom metodom, posebice u podatkovnom prostoru potrebnom za pohranu matrica i rješavanju matričnih jednadžbi. Rešetkasta metoda donosi jednostavnu i modularnu blokovsku izvedbu i eliminira potrebu za eksplicitnim računanjem autokorelacija.

3. Kepstralna analiza

Prema uobičajenom modelu, govorni signal nastaje prolaskom pobudnog signala kroz digitalni filtar koji predstavlja vokalni trakt. U vremenskoj domeni to znači da je govorni signal $s(n)$ konvolucija pobudnog signala $u(n)$ i impulsnog odziva vokalnog trakta $\vartheta(n)$

$$s(n) = u(n) * \vartheta(n). \tag{3.1}$$

Često je poželjno razdvojiti konvoluirane signale radi učinkovitijeg kodiranja ili primjene u postupcima za prepoznavanje, ali to nije jednostavan zadatak. Budući da su signali nelinearno kombinirani, prikaz signala u frekvencijskoj domeni neće biti od koristi i potrebno je pronaći drugi pristup. Jedan od najčešće korištenih je estimacija parametara vokalnog trakta pomoću linearne predikcije, postupak prikazan u prethodnom poglavlju.

U ovom će se poglavlju objasniti kako se korištenjem kepstralne reprezentacije govorni signal može prikazati kao dvije odvojene, linearno kombinirane komponente u kepstru. Svrha takvog prikaza je direktno mjerenje nekih značajki pobudnog signala i vokalnog trakta koje se mogu koristiti u svrhu kodiranja ili, korištenjem prikladnih mjera udaljenosti, prepoznavanja govora. Naglasak će biti na realnom kepstru koji je od veće važnosti u postupku prepoznavanja govora, a kompleksni će se kepstar obraditi bez ulaženja u neke detalje.

3.1. Realni kepstar

Realni kepstar signala $s(n)$ definiran je kao

$$c_s(n) = \mathcal{F}^{-1}\{\log|\mathcal{F}\{s(n)\}|\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S(\omega)|\, e^{j\omega n} d\omega \tag{3.2}$$

gdje je sa $\mathcal{F}$ označena vremenski diskretna Fourierova transformacija (DTFT), a logaritam može biti proizvoljne baze. U nastavku će se pretpostaviti korištenje prirodnog logaritma. Vidljivo je da se uzimanjem apsolutne vrijednosti spektra gubi informacija u fazi, dok u slučaju kompleksnog kepstra koji će biti opisan kasnije ona ostaje očuvana. Iako se gubitkom informacije o fazi gubi i mogućnost rekonstruiranja signala u vremenskoj domeni, kod mjerenja udaljenosti govornih signala dovoljno je poznavati koeficijente realnog kepstra.

Slika 3. Blok dijagram izračuna kepstra govornog signala

Blok dijagram na slici 3 prikazuje proces nalaženja realnog kepstra govornog signala. Postupak računanja kepstra bit će prikazan za slučaj zvučnog govora kod kojeg je $s(n)$ periodički signal i njegov DTFT nije definiran. Međutim, spektar signala može se dobiti generaliziranom Fourierovom transformacijom. Ako se pobudni signal $u(n)$ zapiše pomoću Kroeneckerove delta funkcije označene sa δ_K

$$u(n) = \sum_k \delta_K(n - kP) \tag{3.3}$$

onda je spektar pobudnog signala

$$U(e^{j\omega}) = \mathcal{F}\{u(n)\} = \frac{2\pi}{P} \sum_{k=-\infty}^{\infty} \delta_D\left(\omega - k\frac{2\pi}{P}\right) \tag{3.4}$$

gdje je δ_D Diracova delta funkcija. Ako je $\theta(e^{j\omega})$ Fourierova transformacija funkcije vokalnog trakta $\vartheta(n)$ onda za spektar govornog signala vrijedi:

$$\begin{aligned} S(e^{j\omega}) &= \mathcal{F}\{s(n)\} = \mathcal{F}\{u(n) * \vartheta(n)\} = U(e^{j\omega})\theta(e^{j\omega}) \\ &= \frac{2\pi}{P} \sum_{k=-\infty}^{\infty} \theta\left(k\frac{2\pi}{P}\right) \delta_D\left(\omega - k\frac{2\pi}{P}\right). \end{aligned} \tag{3.5}$$

Gornji izraz predstavlja točan izraz za DTFT modela govornog signala, no pojava Diracove delta funkcije dovest će do velikih teoretskih problema kod daljnjeg računanja. Kako bi se izbjegle Diracove funkcija u spektru, govorni signal $s(n)$ ograničava se simetričnim pravokutnim signalom $w(n)$ duljine L

$$w(n) = \begin{cases} 1, & -\frac{L}{2} \leq n \leq \frac{L}{2} \\ 0, & inače \end{cases} \tag{3.6}$$

$$\tilde{s}(n) = s(n)w(n) \tag{3.7}$$

gdje je L proizvoljan paran broj. DTFT od $w(n)$ je

$$W\left(e^{j\omega}\right) = \mathcal{F}\{w(n)\} = \frac{\sin\left(\frac{\omega(L+1)}{2}\right)}{\sin\left(\frac{\omega}{2}\right)}. \tag{3.8}$$

Budući da je DTFT umnoška jednak konvoluciji pojedinačnih DTFT-ova, spektar ograničenog govornog signala $\tilde{s}(n)$ jednak je konvoluciji spektara govornog signala $s(n)$ i pravokutnog signala $w(n)$. U praksi to znači da se na mjestu svakog Diracovog impulsa u spektru $S\left(e^{j\omega}\right)$ javlja valni oblik spektra $W\left(e^{j\omega}\right)$, ali oblik ovojnice ostaje isti. Amplituda k-tog harmonika spektra $\tilde{S}\left(e^{j\omega}\right)$ iznosi

$$\left|\tilde{S}_k\left(e^{j\omega}\right)\right| = \frac{L+1}{P}\left|\theta\left(k\frac{2\pi}{P}\right)\right| \tag{3.9}$$

za svaki k, a širina glavne "latice" $\frac{2*2\pi}{L+1}$.

Očito je da porastom širine L pravokutnog signala $w(n)$ spektar $\tilde{S}\left(e^{j\omega}\right)$ teži spektru $S\left(e^{j\omega}\right)$. Ovime je riješen problem Diracovih impulsa uz po želji malu pogrešku u spektru signala.

Nadalje će se ograničeni signal i njegov spektar označavati bez tilde, kao $s(n)$, odnosno $S\left(e^{j\omega}\right)$.

Sad kada je poznat spektar govornog signala, logaritmiranjem njegove apsolutne vrijednosti dobiva se

$$\begin{aligned} C_s(e^{j\omega}) &= \log|S(e^{j\omega})| = \log|U(e^{j\omega})\theta(e^{j\omega})| \\ &= \log|U(e^{j\omega})| + \log|\theta(e^{j\omega})| \\ &= C_u(e^{j\omega}) + C_\theta(e^{j\omega}). \end{aligned} \tag{3.10}$$

Ovim se postupkom dobila transformacija govornog signala $C_s(e^{j\omega})$ izražena kao linearna kombinacija komponente ovisne samo o pobudnom signalu i komponente ovisne samo o funkciji vokalnog trakta. Funkcija $C_s(e^{j\omega})$ je periodična s periodom 2π i, promotri li ju se kao signal, moguće je Fourierovom analizom (CTFS) rastaviti u red i tako prikazati u "frekvencijskoj" domeni. Koeficijenti CTFS-a računaju se kao

$$\alpha_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_s(e^{j\omega})e^{-j\omega n}d\omega. \tag{3.11}$$

Po definiciji, realni kepstar signala je

$$c_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_s(e^{j\omega})e^{j\omega n}d\omega. \tag{3.12}$$

Budući da je $C_s(e^{j\omega})$ parna funkcija, gornja dva izraza su istovjetna i realni se kepstar signala može protumačiti kao prikaz "signala" $C_s(e^{j\omega})$ u "frekvencijskoj" domeni. Po konvenciji, kepstralna domena naziva se *kvefrencijska*, "harmonici" se nazivaju *rahmonici*, a "filtriranje" *liftriranje*. Kepstralni koeficijenti $c_s(n)$ u nastavku će biti jednostavnije označavani sa c_n.

3.1.1. Kratkotrajni realni kepstar

U praksi, realni se kepstar računa na kratkim segmentima govora te je potrebno napraviti prijelaz sa dugotrajnog kepstra na kratkotrajni. Prikazom dugotrajnog kepstra iz izraza (3.12) sa eksplicitno izraženim članom $C_s(e^{j\omega})$ dobiva se

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\log\left|\sum_{l=-\infty}^{\infty} s(l)e^{-j\omega l}\right|\right\}e^{j\omega n}d\omega. \tag{3.13}$$

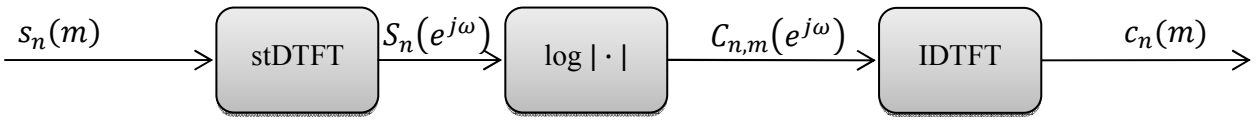
Za segment govornog signala duljine N koji počinje sa indeksom n, $s_n(m), m = 0, \dots, N-1$ umjesto DTFT-a računa se kratkotrajni DTFT (stDTFT)

$$\mathcal{F}\{s_n(m)\} = \sum_{l=n}^{n+N-1} s_n(l)\, e^{-j\omega l} \tag{3.14}$$

iz kojega slijedi izraz za kratkotrajni realni kepstar signala:

$$c_n(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\log\left|\sum_{l=n}^{n+N-1} s_n(l)e^{-j\omega l}\right|\right\}e^{j\omega n}d\omega. \tag{3.15}$$

Proces računanja kratkotrajnog realnog kepstra može se prikazati blok dijagramom na slici 4.



Slika 4. Blok dijagram izračuna kratkotrajnog realnog kepstra

Zanimljivo je uočiti da je na izlazu ostavljen "klasični", a ne kratkotrajni IDTFT. Na taj se način za svaki segment na izlazu dobivaju kepstralni koeficijenti za $m = 0,1, \dots, N-1$. Korištenjem stIDTFT-a obnovila bi se informacija o fazi segmenta m i na izlazu bi se pojavili koeficijenti za $m = n, \dots, n+N-1$, ali bi sve druge informacije o fazi signala i dalje bile izgubljene. Koeficijenti kratkotrajnog kepstra u nastavku će umjesto sa $c_n(m)$ biti označavani c_m uz $m = 0, \dots, N-1$.

3.2. Računanje kepstralnih koeficijenata

Izravno računanje kepstralnih koeficijenata po definicijskoj formuli je vrlo zahtjevan proces. Potrebno je izračunati FFT signala, logaritmirati ga i izračunati integral. U nastavku će

biti opisana dva alternativna načina koji se u praksi gotovo uvijek koriste za računanje kepstra govornog signala.

3.2.1. Računanje kepstra iz LPC koeficijenata

Područje linearne predikcije istraživano je godinama te su razvijeni jaki matematički temelji i učinkoviti algoritmi za računanje LPC koeficijenata. Kepstralni koeficijenti pokazali su se boljom parametrizacijom u području prepoznavanja govora, ali je njihovo računanje znatno složenije. Međutim, pokazalo se da se kepstralni koeficijenti mogu izračunati iz poznatih LPC koeficijenata. Koeficijent c_0 dobiva izravno kao logaritam pojačanja G normaliziranog signala pobude u LPC modelu (odnosno kao logaritam kvadrata pojačanja ako se promatra spektar snage signala)

$$c_0 = \log G \tag{3.16}$$

Ostali se koeficijenti računaju rekurzivno

$$c_m = a_m + \sum_{k=1}^{m} \frac{(m-k)}{m} c_{m-k} a_k\,, \qquad 1 \le m \le p \tag{3.17}$$

$$c_m = \sum_{k=1}^{m} \frac{(m-k)}{m} c_{m-k} a_k\,, \qquad m > p. \tag{3.18}$$

Gornje se relacije dobivaju razvojem logaritma LPC modela u Laurentov red

$$\log \frac{G}{A(z)} = \log G + \sum_{n=1}^{\infty} c_n z^{-n}\,, \tag{3.19}$$

$$-\log\left(1 - \sum_{i=1}^{p} a_i z^{-i}\right) = \sum_{n=1}^{\infty} c_n z^{-n}. \tag{3.20}$$

Nulti koeficijent odmah se može iščitati kao $c_0 = \log G$. Derivacijom gornjeg izraza po z^{-1} dobiva se

$$\frac{1}{1-\sum_{i=1}^{p} a_i z^{-i}} \sum_{i=1}^{p} i a_i z^{1-i} = \sum_{n=1}^{\infty} n c_n z^{1-n}, \tag{3.21}$$

$$\sum_{i=1}^{p} i a_i z^{1-i} = \sum_{n=1}^{\infty} n c_n z^{1-n} - \sum_{i=1}^{p} a_i z^{-i} \sum_{n=1}^{\infty} n c_n z^{1-n}. \tag{3.22}$$

Sada je potrebno izjednačiti članove sa istom potencijom uz z, npr. z^{1-m}. Trivijalno je pronaći članove sume na lijevoj strani izraza i prve sume na desnoj strani izraza koji stoje uz z^{1-m} i oni iznose ma_m i mc_m. Međutim, umnožak suma na desnoj strani potrebno je svesti na malo drugačiji oblik:

$$\sum_{i=1}^{p} a_i z^{-i} \sum_{n=1}^{\infty} n c_n z^{1-n} = \sum_{n=1}^{\infty} n c_n a_1 z^{-n} + \cdots + \sum_{n=1}^{\infty} n c_n a_1 z^{-n-p+1} \tag{3.23}$$

i iz ovakvog se oblika iščitava član sa z^{1-m}

$$\begin{aligned} &(m-1) c_{m-1} a_1 z^{1-m} + \cdots + (m-p) c_{m-p} a_p z^{1-m} = \\ &= z^{1-m} \sum_{k=1}^{p} (m-k) c_{m-k} a_k. \end{aligned} \tag{3.24}$$

Konačno, uvrštavanjem tog izraza u (3.23)

$$a_m m z^{1-m} = c_m m z^{1-m} - z^{1-m} \sum_{k=1}^{p} (m-k) c_{m-k} a_k, \tag{3.25}$$

$$c_m = a_m + \sum_{k=1}^{m} \frac{(m-k)}{m} c_{m-k} a_k. \tag{3.26}$$

Gornja relacija vrijedi za $m \leq p$, dok se za $m > p$ gubi član a_m čime se dolazi do izraza (3.18). Treba napomenuti da kepstralni koeficijenti dobiveni iz LPC koeficijenata više ne predstavljaju kepstar originalnog govornog signal, već kepstar uglađenog modela dobivenog LPC analizom.

3.2.2. Računanje kepstra iz uzoraka govora

Kepstralni se koeficijenti mogu izračunati i izravno iz uzoraka govornog signala. U općem slučaju, *z*-transformacija govornog signala je

$$S(z) = \frac{Gz^r\left[\prod_{k=1}^{o_i}(1-u_k z^{-1})\right]\left[\prod_{k=1}^{o_o}(1-v_k z^{-1})\right]}{\left[\prod_{k=1}^{p_i}(1-x_k z^{-1})\right]\left[\prod_{k=1}^{p_o}(1-y_k z^{-1})\right]}$$

gdje su $(1-u_k z^{-1})$ i $(1-x_k z^{-1})$ nule i polovi unutar jedinične kružnice, a $(1-v_k z^{-1})$ i $(1-y_k z^{-1})$ nule i polovi izvan jedinične kružnice. Korištenjem LPC modela radi se o minimalno faznom signalu kojem se svi polovi i nule nalaze unutar jedinične kružnice. U tom se slučaju pokazuje [1] da se kepstralni koeficijenti mogu izračunati prema izrazima

$$c_0 = \log G$$

$$c_n = \sum_{k=1}^{p_i}\frac{x_k^n}{n} - \sum_{k=1}^{o_i}\frac{u_k^n}{n}, \qquad n > 0$$

dok su koeficijenti za $n < 0$ jednaki nuli. U *all-pole* modelu kakav se često promatra koeficijenti u_k su jednaki nuli.

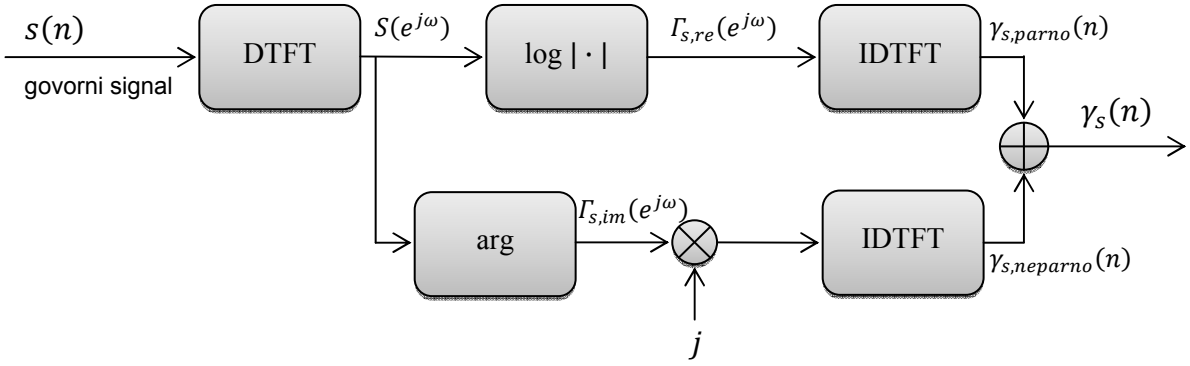
3.3. Kompleksni kepstar

U poglavlju o realnom kepstru već je spomenuto da se gubitkom informacije o fazi gubi i mogućnost točnog rekonstruiranja signala u vremenskoj domeni. U mjerama udaljenosti govora to često ne predstavlja problem jer koeficijenti realnog kepstra pružaju potrebnu informaciju. Kompleksni kepstar čuvanjem informacije o fazi signala nudi mogućnost transformacije nelinearno kombiniranih signala u kepstralnu domenu, liftriranja i izdvajanja kepstra željenog signala te povratka u vremensku domenu. U području obrade govora to znači da bi se u kepstru mogle odvojiti pobuda i prijenosna funkcija vokalnog trakta (što je moguće i u realnom kepstru), te zatim nezavisno prikazati u frekvencijskoj ili u vremenskoj domeni.

Definicija kompleksnog kepstra (označenog sa γ_s) gotovo je istovjetna definiciji realnog kepstra. Jedina je razlika računanje logaritma cijelog spektra (uključujući i fazu)

$$\gamma_s(n) = \mathcal{F}^{-1}\{\log \mathcal{F}\{s(n)\}\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log S(\omega)\, e^{j\omega n} d\omega, \tag{3.27}$$

uz logaritam kompleksnog broja z definiran kao $\log z = \log|z| + j \arg z$.



Slika 5. Blok dijagram izračuna kompleksnog kepstra govornog signala

Na blok dijagramu na slici 5 prikazan je proces računanja kompleksnog kepstra. Komponenta dobivena iz amplitudnog dijela, $\log\left|S\left(e^{j\omega}\right)\right|$, realna je, parna i jednaka realnom kepstru

$$\gamma_{s,parno}(n) = c_s(n),$$

dok je komponenta dobivena iz faznog dijela također realna, ali neparna, i predstavlja razliku kompleksnog i realnog kepstra.

4. Mjere udaljenosti govornog signala

Objektivno određivanje sličnosti između dvaju govornih signala nije ni približno jednostavno kao što na prvi pogled djeluje. Postavlja se pitanje kako matematički prikazati razliku između govora, a da ona odgovara subjektivnom dojmu slušatelja – za dva govorna signala koja ljudskom uhu zvuče slično, rezultat matematičkog uspoređivanja treba biti malen, a kako subjektivna razlika između signala raste, rezultat se treba povećavati. Za tu su potrebu proučavane i razvijane mjere udaljenosti dvaju govornih signala. Većina njih kao parametre govora koje uspoređuje koristi kepstralne koeficijente ili LPC koeficijente, opisane u prethodnim poglavljima.

Osim u prepoznavanju, često je i prilikom kodiranja govornog signala potrebno objektivno odrediti sličnost originalnog i kodiranog signala, a odabir metode mjerenja sličnosti ovisi o vrsti kodera. Koderi valnog oblika signala kodiraju signal na način da što bolje sačuvaju njegov valni oblik, te se kod njih koriste mjere odnosa signala i šuma (*SNR ratio measures*), koje u ovom radu neće biti od interesa. Druga vrsta kodera su *vocoderi* koji kodiraju signal sa zahtjevom da ljudskom uhu zvuči što sličnije originalu. U tom se postupku obično čuva samo informacija o amplitudi signala, dok se, prema hipotezi da je ljudsko uho neosjetljivo na kratkotrajnu fazu, faza signala odbaciju. Time se može postići da kodirani signal zvuči veoma slično originalnom, ali uz veliku razliku u valnom obliku. Korištenje *SNR* mjera tada nema smisla i kod takvih se kodera objektivna sličnost određuje mjerama udaljenosti govornog signala.

U nastavku će se opisati neke od najčešće korištenih mjera udaljenosti govornog signala, usporediti njihova učinkovitost te računalna složenost i zahtjevi za podatkovnim prostorom potrebnim za njihovu implementaciju.

4.1. Metrika – matematički pogled na funkcije udaljenosti

Neka je $X \neq \emptyset$ neprazan skup. Metrika ili funkcija udaljenosti $d : X \times X \to \mathbb{R}$ na skupu X definirana je kao preslikavanje sa Kartezijevog produkta $X \times X$ u skup realnih brojeva $\mathbb{R}$ za koje vrijede sljedeća svojstva:

a) $d(a,b) \geq 0,\ \forall a,b \in X$

b) $d(a,b) = 0 \Leftrightarrow a = b$

c) $d(a,b) = d(b,a),\ \forall a,b \in X$

d) $d(a,b) \leq d(a,c) + d(c,b),\ \forall a,b,c \in X$

Svojstva a) i b) čine aksiom pozitivnosti, svojstvo c) simetričnosti, a svojstvo d) nejednakosti trokuta. Može se pokazati i da a) i c) slijede iz b) i d). [6]

Kod mjerenja udaljenosti govornih signala važno je postići ranije spomenutu korelaciju sa subjektivnom sličnošću. Pokazalo se da je veoma teško udovoljiti i matematičkoj definiciji funkcije udaljenosti i subjektivnom dojmu slušatelja. Neke mjere udaljenosti stoga neće biti funkcije udaljenosti po strogo matematičkog definiciji.

4.2. Utjecaj promjena u spektru na percepciju govora

Mjere udaljenosti govornih signala zasnivaju se na mjerenju razlike spektralnih karakteristika za koje se pokazalo da utječu na ljudsku percepciju zvuka. Korisno je razlučiti koje promjene u spektru utječu, a koje ne na dojam slušatelja o govornom signalu, odnosno koje su promjene u spektru fonetički značajne.

Označi li se sa $S_1(\omega)$ i $S_2(\omega)$ spektri signala $s_1(n)$ i $s_2(n)$, spektralne promjene koje ne utječu značajno na subjektivni doživljaj govora su:

- spektralni nagib - $S_2(\omega) = S_1(\omega) \cdot \omega^{\alpha}$;
- visokopropusno filtriranje - $S_2(\omega) = S_1(\omega)\left|H_H\left(e^{j\omega}\right)\right|^2$, gdje je $H_H\left(e^{j\omega}\right)$ visokopropusni filtar s donjom graničnom frekvencijom ispod frekvencije prvog formanta;
- niskopropusno filtriranje - $S_2(\omega) = S_1(\omega)\left|H_L\left(e^{j\omega}\right)\right|^2$, gdje je $H_L\left(e^{j\omega}\right)$ niskopropusni filtar s gornjom graničnom frekvencijom iznad frekvencije trećeg ili četvrtog formanta;
- filtriranje pojasnom branom - $S_2(\omega) = S_1(\omega)\left|H_N\left(e^{j\omega}\right)\right|^2$, gdje je $H_N\left(e^{j\omega}\right)$ filtar s konstantnom vrijednošću osim za vrlo uzak pojas frekvencija gdje je je signal jako prigušen.

Mjera udaljenosti $d(S_1, S_2)$ trebala bi za gornje slučajeve rezultirati malom udaljenošću.

Spektralne promjene koje dovode do promjene u percepciji govora su:

- primjetne razlike u frekvencijama formanata signala $s_1(n)$ i $s_2(n)$;
- primjetne razlike u širinama pojasa formanata signala $s_1(n)$ i $s_2(n)$.

Za gornje slučajeve $d(S_1, S_2)$ trebala bi rezultirati velikom udaljenošću. Postavlja se pitanje kolike su te "primjetne razlike" u frekvencijama i širinama pojasa formanata koje rezultiraju različitom percepcijom dvaju signala. Koristeći računalno sintetiziran govor, empirički se pokazalo da dolazi do opazive razlike u zvuku kod promjene osnovne frekvencije govora 0,3-0,5%, frekvencije formanta 3-5%, širine pojasa formanta 20-40% ili promjene intenziteta govora od 1,5dB. Vidljivo je da je potrebna promjena frekvencije formanta jedan, a promjena širine pojasa formanta dva reda veličine veća od promjene osnovne frekvencije i na prvi pogled djeluje da će osnovna frekvencija imati najveću ulogu u mjerenju udaljenosti. To ipak nije slučaj jer, iako se promjenom osnovne frekvencije mjenja naša percepcija zvuka, fonetski ne dolazi do značajnih promjena. Stoga promjena osnovne frekvencije nije među karakteristikama značajnima za mjere udaljenosti, a frekvencija i širina pojasa formanata jesu. To je i jedan od razloga zašto se kod mjerenja udaljenosti umjesto spektra govora dobivenog FFT-om koristi model prijenosne funkcije vokalnog trakta dobiven LPC analizom, koji odbacuje informaciju o pobudnom signalu (a samim time i o osnovnoj frekvenciji), ali uključuje informacije o formantima i intenzitetu.

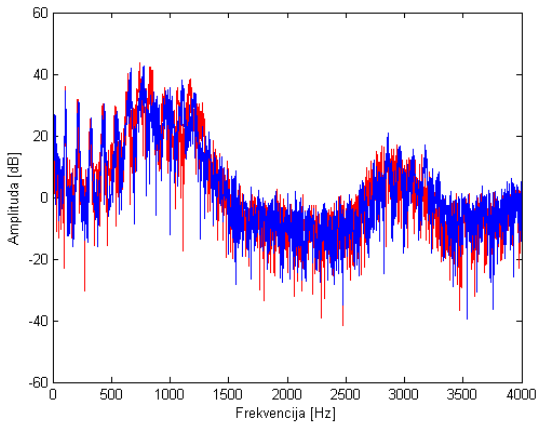
4.3. Pregled mjera udaljenosti

Ljudska percepcija glasnoće zvuka je približno logaritamska pa je i logično da se u većini mjera udaljenosti koristi logaritam spektra. Može se već i naslutiti da će zbog toga kepstar imati veliku ulogu. Korištenje logaritma spektra dovodi i do toga da je praktički svejedno promatramo li amplitudni spektar ili spektar snage signala budući da će vrijediti $S_{am}(\omega) = KS_{sn}(\omega)$, gdje su $S_{am}(\omega)$ i $S_{sn}(\omega)$ amplitudni spektar i spektar snage. Ipak, zbog nekih veza između spektra snage i autokorelacija signala koje će kasnije biti korištene, u nastavku će se koristiti spektar snage označen skraćeno sa $S(\omega)$.

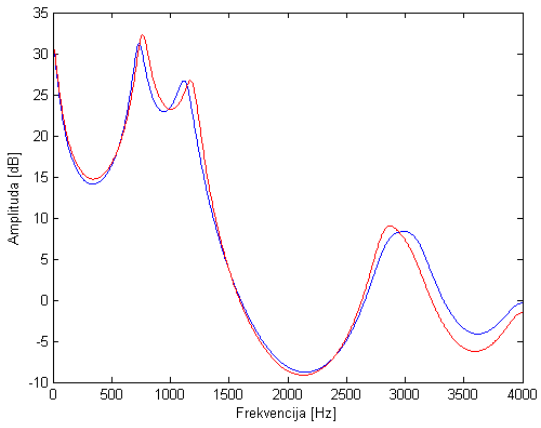
4.3.1. Spektralni model govornog signala u mjerama udaljenosti

Mjere udaljenosti bit će opisane na primjeru dvaju spektara $S(\omega)$ i $S'(\omega)$ govornih signala. Spektri se mogu izračunati izravno primjenom brze Fourierove transformacije (FFT) na govorne signale $s(n)$ i $s'(n)$ čime se dobiva i informacija o osnovnoj frekvenciji govora. Razlika u osnovnoj frekvenciji fonetski nije značajna, ali bi mogla rezultirati velikom udaljenošću. Iz tog se razloga kao spektralna reprezentacija govora u pravilu koristi spektar tzv. *all-pole* modela oblika $\sigma/A(z)$, gdje $A(z)$ polinom p-tog stupnja sa koeficijentima a_i izračunatima postupkom LPC analize. Međutim, kao što je ranije napomenuto koristit će se spektar snage govornog signala oblika $\sigma^2/[A(z)]^2$.

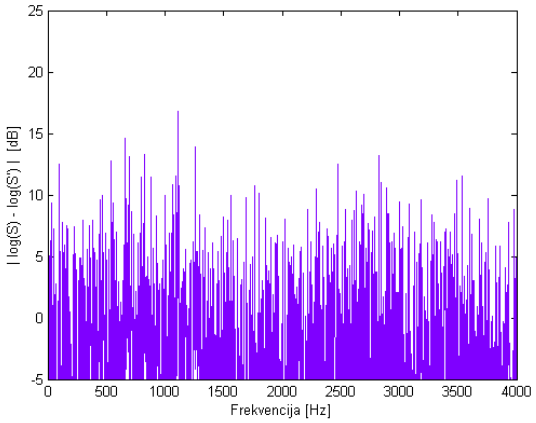
Na slici 6 prikazan je odnos FFT spektara dvaju signala glasa "A", a na slici 7 odnos spektara LPC modela istih dvaju signala. Vidi se da spektri LPC modela približno odgovaraju uglađenim spektrima signala. Na slikama 8 i 9 prikazana je amplituda razlike logaritama spektara FFT-a signala, odnosno amplituda razlike logaritama spektara LPC modela signala. Očito je da kod FFT-a visoke frekvencije u spektru mogu uzrokovati velike razlike između dvaju iznimno sličnih spektara, dok su kod LPC modela razlike mnogo manje i javljaju se većinom na frekvencijama formanata, što odgovara zahtjevima kod mjerenja udaljenosti govora.



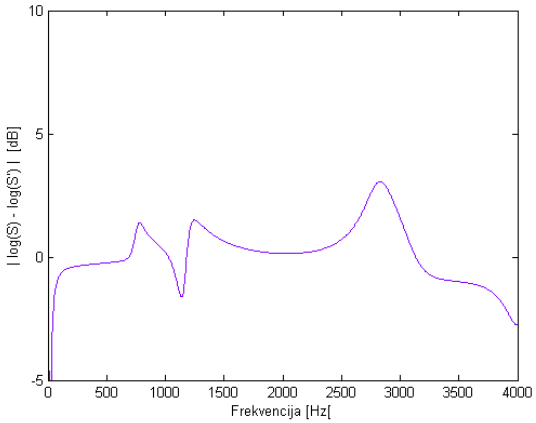
Slika 6. FFT dvaju signala glasa "A"



Slika 7. Spektar LPC *all-pole* modela dvaju signala glasa "A"

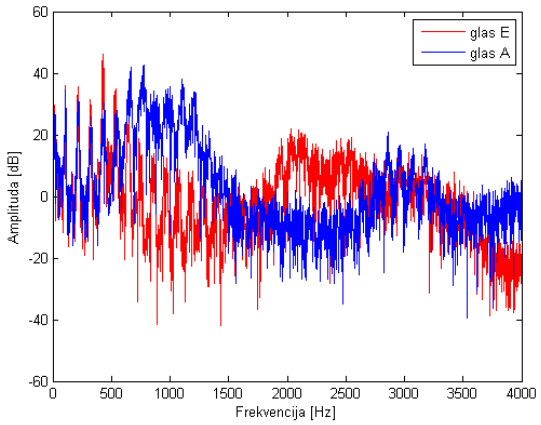


Slika 8. Razlika logaritama amplituda FFT-a dvaju signala glasa "A"

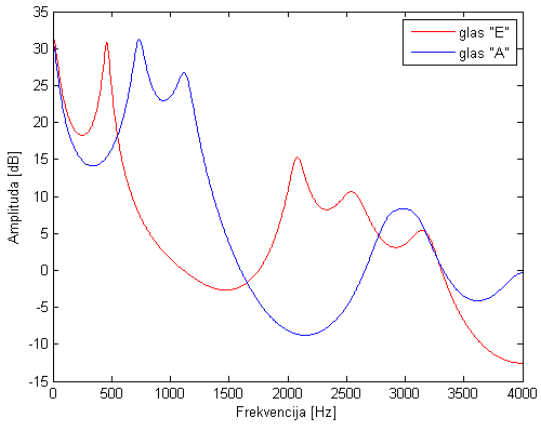


Slika 9. Razlika logaritama amplituda LPC modela dvaju signala glasa "A"

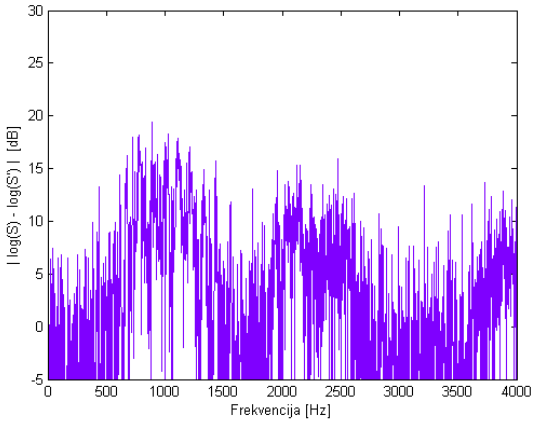
Na slici 10 prikazan je FFT spektar glasova "A" i "E", a na slici 11 spektar LPC modela istih glasova. Ponovo se može uočiti odlično preklapanje LPC modela i spektra signala, uz eliminaciju visokih frekvencija u spektru. Razlike logaritama spektara FFT-a te spektara LPC modela prikazane su na slikama 12 i 13. Usporedbom sa slikama 8 i 9, gdje su prikazane razlike dvaju signala istog glasa "A", vidi se da se razlika spektara LPC modela mnogo više povećala od razlike spektara FFT-a signala.



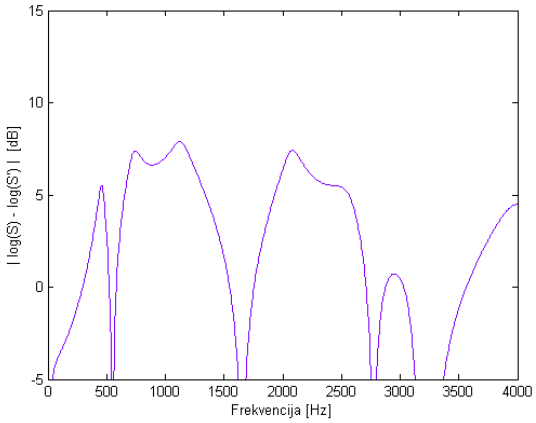
Slika 10. FFT signala glasova "A" i "E"



Slika 11. Spektar LPC *all-pole* modela signala glasova "A" i "E"



Slika 12. Razlika logaritama amplituda FFT-a signala glasova "A" i "E"



Slika 13. Razlika logaritama amplituda LPC modela signala glasova "A" i "E"

4.3.2. Veza spektra snage i autokorelacija govornog signala

Prije pregleda samih mjera udaljenosti nužno je proučiti neke odnose spektra snage i autokorelacija signala.

Neka je $S(\omega)$ spektralna gustoća govornog signala $s(i)$, sa normaliziranom frekvencijom ω u rasponu od $-\pi$ do π i Fourierovim koeficijentima $r(n)$ koji definiraju autokorelacije

$$S(\omega) = \sum_{-\infty}^{\infty} r(n) e^{-j\omega n}, \tag{4.1}$$

$$r(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) e^{j\omega n} d\omega. \tag{4.2}$$

Funkcija autokorelacije govornog signala je

$$r_E(n) = E\{s(i)s(i+n)\}. \tag{4.3}$$

Ako se uzme $r(n) = r_E(n)$, tada $S(\omega)$ predstavlja spektralnu gustoću snage signala. Međutim, govorni se signal zbog vremenske promjenjivosti promatra u kratkim segmentima duljine N. Kratkotrajna autokorelacija takvog segmenta definirana je kao

$$r_N(n) = \sum_{i=0}^{N-|n|-1} s(i)s(i+|n|), \qquad n = 0,1,\dots,N-1. \tag{4.4}$$

Odabirom $r(n) = r_E(n)$ za $n < N$ te $r(n) = 0$ za $n \geq N$, $S(\omega)$ postaje spektralna gustoća energije. Uz pretpostavku ergodičnosti govornog signala (koja će se uvijek smatrati istinitom), kratkotrajna autokorelacija $r_N(n)$, normalizirana duljinom segmenta N, teži autokorelaciji $r_E(n)$ te će se nadalje $S(\omega)$ promatrati kao spektralna gustoća snage, a $r(n)$ kao kratkotrajne autokorelacije.

Također, neka je spektar signala predstavljen *all-pole* modelom oblika $\sigma/A(z)$, gdje je $A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$ prijenosna funkcija pogreške predikcije prediktora reda p. Energija pogreške predikcije tada ovisi o odabiru koeficijenata a_i i jednaka je

$$E(\boldsymbol{a}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|1 - \sum_{i=1}^{p} a_i e^{-j\omega i}\right|^2 S(\omega)d\omega = \boldsymbol{a}^t \boldsymbol{R}_p \boldsymbol{a}. \tag{4.5}$$

U gornjem izrazu $\boldsymbol{a}^t$ označava vektor koeficijenata a_i, a $\boldsymbol{R}_p$ je Toeplitzova matrica autokorelacija dimenzije $(p+1) \times (p+1)$

$$\boldsymbol{R}_p = \begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(p) \\ r(1) & r(0) & r(1) & \cdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ r(p) & r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}. \tag{4.6}$$

U poglavlju o linearnom prediktivnom kodiranju objašnjeno je kako se optimalni LPC koeficijenti $\boldsymbol{a}_p$ dobivaju minimizacijom pogreške predikcije

$$\boldsymbol{a}_p = \min_{\boldsymbol{a}} E(\boldsymbol{a}) = \min_{\boldsymbol{a}} \boldsymbol{a}^t \boldsymbol{R}_p \boldsymbol{a}. \tag{4.7}$$

Minimum pogreške predikcije tada iznosi $E(\boldsymbol{a}_p) = \sigma_p^2$ i optimalni spektar snage je

$$S(\omega) = \frac{\sigma_p^2}{\left|A_p(e^{j\omega})\right|^2}. \tag{4.8}$$

Za σ_p^2 se pokazuje [2] da se može izračunati iz rekurzivnog izraza

$$\sigma_p^2 = \frac{|\boldsymbol{R}_p|}{|\boldsymbol{R}_{p-1}|}. \tag{4.9}$$

U pregledu mjera udaljenosti važna će biti i pogreška koraka predikcije

$$\sigma_\infty^2 = \lim_{p\to\infty} \sigma_p^2 = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log S(\omega) d\omega\right) \tag{4.10}$$

te svojstvo da je očekivanje logaritma spektra minimalno faznog *all-pole* modela jednako nuli,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{1}{|A(e^{j\omega})|^2} d\omega = 0. \tag{4.11}$$

4.3.3. Udaljenost logaritama spektara

Uzevši u obzir ranije spomenutu spoznaju da je ljudska percepcija zvuka logaritamska, prirodno je definirati mjeru udaljenosti kao ukupnu akumuliranu razliku logaritama spektara. Skup takvih L_p normi dan je izrazom

$$d_p(S, S') = \sqrt[p]{\frac{1}{2\pi}\int_{-\pi}^{\pi} |\log S(\omega) - \log S'(\omega)|^p \, d\omega}. \tag{4.12}$$

Ovako definirane mjere udaljenosti zadovoljavaju svojstva pozitivnosti, simetričnosti i nejednakosti trokuta te stoga spadaju u metrike.

4.3.4. Udaljenost kepstara

Sukladno razmatranjima u poglavlju o kepstralnoj analizi, logaritam spektra signala može se izraziti preko kepstralnih koeficijenata kao

$$\log \mathrm{S}(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-j\omega n}. \tag{4.13}$$

Korištenjem gornje relacije udaljenost logaritama spektara d_2 može se napisati kao

$$d_2(S, S') = \sqrt{\sum_{n=-\infty}^{\infty} (c_n - c'_n)^2}\,. \tag{4.14}$$

U poglavlju o kepstru pokazalo se da su u slučaju minimalno faznog modela govornog signala svi negativni kepstralni koeficijenti jednaki nuli dok su pozitivni asimptotski ograničeni. Iako postoji beskonačno mnogo kepstralnih koeficijenata, iz veze kepstralnih i LPC koeficijenata vidi se da prvih p koeficijenata jednoznačno određuje minimalno fazni *all-pole* filtar. Uz korištenje $L \geq p$ kepstralnih koeficijenata, skraćenjem izraza (4.14) na koeficijente $1 \leq n \leq L$ dobiva se mjera udaljenosti kepstara

$$d_c(L) = \sqrt{\sum_{n=1}^{L} (c_n - c'_n)^2}\,. \tag{4.15}$$

Ova mjera već za male L ($p \leq L \leq 2p$) daje rezultate gotovo identične izrazu (4.14) (odnosno mjeri udaljenosti logaritama spektara ako se kepstar računa po definiciji, a ne iz LPC koeficijenata) uz mnogo jednostavniji i brži račun. Korištenje kepstra u mjeri udaljenosti donosi još prednosti opisane u idućem poglavlju, u kojem će se razmatrati korištenje težinskih faktora nad kepstralnim koeficijentima. Mjera udaljenosti kepstara također je funkcija udaljenosti u matematičkom smislu.

4.3.5. Težinska udaljenost kepstara

Za kepstralne se koeficijente, izuzev c_0, može pokazati da imaju očekivanja nula i varijance obrnuto proporcionalne indeksu koeficijenta

$$E(c_n) = 0, \qquad \forall n \neq 0$$

$$\sigma^2(c_n) \sim \frac{1}{n^2}, \qquad \forall n \neq 0.$$

Ta se činjenica može iskoristiti za definiranje nove mjere udaljenosti, dobivene normalizacijom doprinosa svakog kepstralnog koeficijenta s obzirom na njegovu varijancu. To se postiže množenjem koeficijenata s faktorom n^2

$$d = \sqrt{\sum_{n=0}^{\infty} n^2 (c_n - c'_n)^2} = \sqrt{\sum_{n=0}^{\infty} (nc_n - nc'_n)^2}. \tag{4.16}$$

Niski kepstralni koeficijenti vrlo su ovisni o govorniku i nekim drugim čimbenicima koji se manifestiraju spektralnim nagibom. Ideja mjera udaljenosti govora je da ne ovise o govorniku, već samo o fonetskom sadržaju, a već je objašnjeno da spektralni nagib spada u spektralne promjene koje fonetski ne utječu na govor. Uzevši to u obzir, gore definirana mjera udaljenosti predstavlja opravdan izbor budući da niže kepstralne koeficijente uzima sa manjim težinskim faktorima.

Treba proučiti i kakva je uloga visokih kepstralnih koeficijenata govora i koliko su oni važni u definiranju dobre mjere udaljenosti. Neka je $s_i(n)$ izvorni govorni signal, a $s_l(n)$ LPC model tog govornog signala dobiven konvolucijom pobudnog signala i *all-pole* modela vokalnog trakta. Uspoređivanjem varijance kepstralnih koeficijenata $\sigma_i^2(n)$ izvornog signala i varijance kepstralnih koeficijenata $\sigma_l^2(n)$ LPC modela pokazuje se da potonje rastu mnogo brže, odnosno omjer $\sigma_l^2(n)/\sigma_i^2(n)$ raste s n. To je posljedica pogrešaka LPC analize koja se nasljeđuje u kepstru i očito je opravdano u mjeri udaljenosti zanemariti visoke kepstralne koeficijente. Također, umjesto težinskog faktora n^2 prikladno je odabrati neku drugu težinsku funkciju $w(n)$ koja će prigušiti i niske i visoke kepstralne koeficijente, a naglasiti srednje. Jedan mogući odabir za težinsku funkciju je

$$w(n) = \begin{cases} 1 + h \sin\left(\frac{n\pi}{L}\right), & n = 1, \dots, L \\ 0, & inače \end{cases} \tag{4.17}$$

gdje je faktor h uobičajeno jednak $L/2$. Kepstralni koeficijenti pomnoženi sa ovako definiranom težinskom funkcijom rezultirati će uglađenim LPC spektrom. Formanti će i dalje biti jasno izraženi, ali neće se pojaviti "šiljci", već uglađena nadvišenja. Time je definirana težinska mjera udaljenosti kepstara

$$d_{wc} = \sqrt{\sum_{n=1}^{L} [w(n)c_n - w(n)c'_n]^2}. \tag{4.18}$$

Težinska mjera udaljenosti kepstara također spada u matematičke funkcije udaljenosti.

4.3.6. Itakura-Saito mjera udaljenosti

Fumitada Itakura i Shuzo Saito predložili su mjeru udaljenosti definiranu kao

$$d_{IS}(S, S') = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[e^{V(\omega)} - V(\omega) - 1\right] d\omega, \tag{4.19}$$

gdje je $V(\omega)$ razlika logaritama spektara signala S i S', $V(\omega) = \log S(\omega) - \log S'(\omega)$. Gornji se izraz sređivanjem i korištenjem (4.10) može zapisati i preko pogrešaka koraka predikcije

$$d_{IS}(S, S') = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{S(\omega)}{S'(\omega)} d\omega - \log \frac{\sigma_\infty^2}{\sigma_\infty'^2} - 1. \tag{4.20}$$

Dva značajna svojstva Itakura-Saito mjere udaljenosti su asimetričnost i aditivnost. Svojstvo aditivnosti je granični slučaj svojstva nejednakosti trokuta opisanog u poglavlju o metrikama, i pokazuje da je udaljenost dvaju signala govora *A* i *B* jednaka zbroju udaljenosti od *A* do *C* i od *C* do *B*

$$d_{IS}(A, B) = d_{IS}(A, C) + d_{IS}(C, B).$$

Asimetričnost je razlog što Itakura-Saito mjera udaljenosti nije u strogo matematičkom smislu funkcija udaljenosti. Odmah je vidljiva iz izraza (4.19), gdje je podintegralna funkcija za $V(\omega) \gg 1$ približno jednaka e^V, a za $V(\omega) \ll 1$ je približno jednaka $-V$. Postoje neka opravdanja proizašla iz subjektivnih testova za korištenje asimetrične mjere udaljenosti. Pokazalo se da ljudsko uho lakše prepoznaje šum u tonu, nego ton u šumu, i Itakura-Saito mjera udaljenosti rezultira većom udaljenošću kad se uspoređuje ton sa šumom nego obratno. Iz Itakura-Saito mjere jednostavno je definirati simetrične mjere kao

$$d_x^{(m)} = \frac{1}{2} \sqrt[m]{[d_{IS}(S, S')]^m + [d_{IS}(S', S)]^m}. \tag{4.21}$$

Takva mjera za $m = 1$ opisana je u nastavku i naziva se COSH mjera udaljenosti.

4.3.7. COSH mjera udaljenosti

Uzme li se simetrična mjera iz izraza (4.21) uz faktor $m = 1$

$$d_x^{(1)} = \frac{1}{2}[d_{IS}(S, S') + d_{IS}(S', S)] \tag{4.22}$$

uvrštavanjam u izraz (4.19) dobije se

$$d_x^{(1)} = \frac{1}{2}\int_{-\pi}^{\pi}\left[e^{V(\omega)} - V(\omega) - 1 + e^{-V(\omega)} + V(\omega) - 1\right]d\omega. \tag{4.23}$$

U gornjem je izrazu lako prepoznati kosinuse hiperbolne pa se takva mjera naziva COSH mjera udaljenosti

$$d_{COSH}(S, S') = \int_{-\pi}^{\pi}\{\cosh[V(\omega)] - 1\}d\omega. \tag{4.24}$$

Za male udaljenosti COSH udaljenost praktički je jednaka dvostrukoj udaljenosti logaritama spektara. Budući da je ostvareno svojstvo simetričnosti, COSH mjera je funkcija udaljenosti i u matematičkom smislu.

4.3.8. Itakurina mjera udaljenosti

Itakura-Saito udaljenost između optimalnog $\sigma_p^2/\left|A_p\right|^2$ te proizvoljnog $\sigma^2/|A|^2$ *all-pole* LPC modela spektra jednaka je

$$d_{IS}\left(\frac{\sigma_p^2}{\left|A_p\right|^2}, \frac{\sigma^2}{|A|^2}\right) = \frac{\sigma_p^2}{\sigma^2}\frac{1}{2\pi}\int_{-.\pi}^{\pi}\frac{|A|^2}{\left|A_p\right|^2}d\omega - \log\frac{\sigma_p^2}{\sigma^2} - 1. \tag{4.25}$$

Neka je pojačanje σ^2 jednako upravo energiji pogreške predikcije E definiranu u (4.5)

$$\sigma^2 = E = \frac{1}{2\pi}\int_{-\pi}^{\pi}S(\omega)\left|A\left(e^{j\omega}\right)\right|^2 d\omega. \tag{4.26}$$

Budući da je prvih $(p+1)$ autokorelacija spektra $S(\omega)$ i optimalnog LPC modela $\sigma_p^2/|A_p|^2$ jednako vrijedi

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)\left|A\left(e^{j\omega}\right)\right|^2 d\omega = \frac{1}{2\pi}\sigma_p^2 \int_{-\pi}^{\pi} \frac{\left|A\left(e^{j\omega}\right)\right|^2}{\left|A_p(e^{j\omega})\right|^2} d\omega \tag{4.27}$$

te iz (4.27), (4.28), (4.29) i (4.5) slijedi

$$d_{IS}\left(\frac{\sigma_p^2}{|A_p|^2}, \frac{E^2}{|A|^2}\right) = \log\frac{E}{\sigma_p^2} = \log\frac{\boldsymbol{a}^t \boldsymbol{R}_p \boldsymbol{a}}{\sigma_p^2}. \tag{4.28}$$

Gornji izraz predstavlja jedan oblik Itakurine mjere udaljenosti i može se shvatiti kao mjera kvalitete nekog LPC modela u odnosu na optimalni model. Budući da je $\boldsymbol{R}_p/\sigma_p^2$ normalizirana matrica autokorelacija s obzirom na pojačanje σ_p^2, Itakurina se mjera udaljenosti može zapisati i u obliku

$$d_I\left(\frac{1}{|A_1|^2}, \frac{1}{|A_2|^2}\right) = \log\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\left|A_2\left(e^{j\omega}\right)\right|^2}{|A_1(e^{j\omega})|^2} d\omega\right\} = \log\frac{\boldsymbol{a}_2^t \boldsymbol{R}_p \boldsymbol{a}_2}{\boldsymbol{a}_1^t \boldsymbol{R}_p \boldsymbol{a}_1}, \tag{4.29}$$

gdje su $1/|A_1|^2$ i $1/|A_2|^2$ LPC modeli spektara dvaju govornih signala, a $\boldsymbol{a}_1^t$ i $\boldsymbol{a}_2^t$ pripadajući vektori LPC koeficijenata. Iz gornjeg izraza se vidi da Itakurina mjera udaljenosti, osim što je asimetrična, ne uzima u obzir faktor pojačanja. O subjektivnoj opravdanosti asimetričnosti već je bilo govora u Itakura-Saito mjeri udaljenosti. Intenzitet govora (uz pretpostavku da je iznad određenog praga) vrlo malo utječe na razumljivost govora pa je zanemarivanje faktora pojačanja djelomično opravdano. Ni Itakurina mjera udaljenosti zbog asimetričnosti ne spada u matematičke funkcije udaljenosti.

4.3.9. Mjera udaljenosti temeljena na omjeru vjerodostojnosti

Mjera udaljenosti temeljena na omjeru vjerodostojnosti proizlazi iz Itakura-Saito mjere udaljenosti te je veoma slična Itakurinoj mjeri udaljenosti. Definirana je kao

$$d_{LR}\left(\frac{1}{|A_1|^2}, \frac{1}{|A_2|^2}\right) = d_{IS}\left(\frac{1}{|A_1|^2}, \frac{1}{|A_2|^2}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\left|A_2\left(e^{j\omega}\right)\right|^2}{|A_1(e^{j\omega})|^2} d\omega - 1 \tag{4.30}$$

i vidi se da je također asimetrična te da zanemaruje faktor pojačanja. Za male udaljenosti gotovo je jednaka Itakurinoj mjeri udaljenosti. Zbog asimetričnosti ni ona ne spada u funkcije udaljenosti.

4.4. Mjerenje udaljenosti samoglasnika

Kako bi rezultati bili što vjerodostojniji, mjere udaljenosti bit će testirane na samoglasnicima. Samoglasnici su stacionarni skoro cijelim trajanjem što omogućuje da se analiza provodi na cijelom glasu, bez djeljenja na segmente. Na taj se način eliminira potreba za vremenskim poravnavanjem promatranih glasova o kojem bi uvelike ovisili rezultati.

4.4.1. Način testiranja mjera udaljenosti i subjektivne ocjene

U svim je mjerama udaljenosti korišten linearni prediktor 12. reda, $p = 12$. Za svaku će se mjeru udaljenosti provesti tri testa. U prvom će se testu mjeriti udaljenosti između samoglasnika koje je izgovorio isti muški govornik, u drugom između samoglasnika koje su izgovorila dva različita muška govornika, a u trećem između samoglasnika muškog i ženskog govornika. U drugom i trećem testu se javlja dodatnih deset mogućih kombinacija glasova i govornika koji se neće razmatrati. Rezultati svakog testa će biti dobiveni kao aritmetička sredina mjerenja iz više različitih snimaka glasova. Takvim usrednjavanjem rezultata želi se smanjiti utjecaj anomalija u pojedinim snimkama na konačan rezultat. Odnosi apsolutnih iznosa udaljenosti između različitih mjera nemaju nikakvo značenje te će stoga rezultati za svaku mjeru biti normirani na najkraću udaljenost. Kepstralne udaljenosti testirat će se uz kepstar izračunat po definiciji te uz kepstralne koeficijente dobivene iz LPC koeficijenata. Za asimetrične mjere udaljenosti prikazat će se udaljenosti u oba smjera, osim za slučaj istih glasova u prvom testu gdje su udaljenosti gotovo jednake.

Prema subjektivnom dojmu nekoliko slušatelja određeni su očekivani rezultati. Isti samoglasnici trebali bi rezultirati najmanjom udaljenošću, posebice u prvom testu s istim govornikom. Isti glasovi različitih govornika (drugi i treći test) trebeli bi rezultirati manjom udaljenošću od različitih glasova istog govornika. Parovi glasova (E,I) i (O,U) subjektivno su sličniji od ostalih parova glasova te bi njihova udaljenost trebala biti veća nego udaljenost parova

istih glasova, ali manja od udaljenosti ostalih parova različitih glasova. Svi su slušatelji rekli da su razlike između glasova pojedinog para simetrične.

4.4.2. Rezultati testova

Udaljenost logaritama spektara

Rezultati testova udaljenosti logaritama spektara prikazani su u tablici 2. U prvom testu udaljenosti između istih glasova su nekoliko puta manje od udaljenosti različitih glasova. Udaljenost $d_2(O,U)$ je manja od ostalih parova glasova, dok je $d_2(E,I)$ nešto veća.

U drugom testu, sa dva muška govornika, udaljenosti istih glasova su također vidljivo manje od udaljenosti različitih glasova, a bitno je primjetiti i da su manje od udaljenosti različitih glasova istog govornika. Udaljenost $d_2(O,U)$ među različitim glasovima, a udaljenost $d_2(I,O)$ je, kao i u prvom testu, manja od $d_2(E,I)$, što ne odgovara subjektivnom dojmu slušatelja.

Treći test daje gotovo iste rezultate kao i drugi, uz iznimku udaljenosti $d_2(I,O)$, koja je sada nešto veća.

Tablica 2. Udaljenosti samoglasnika korištenjem mjere udaljenosti logaritama spektara

$d_2(S,S')$	A,A	E,E	I,I	O,O	U,U	A,E	A,I	A,O	A,U	E,I	E,O	E,U	I,O	I,U	O,U
isti govornik	1,46	1,35	**1**	1,44	1,26	5,11	6,98	5,45	7,65	5,14	5,61	7,37	4,80	5,12	4,21
dva muška govornika	3,03	3,4	2,85	3,31	2,01	6,78	8,32	7,09	9,1	6,7	7,79	9,26	5,67	6,16	5,14
dva govonika, muški i ženski	3,24	3,17	3,48	3,52	1,98	4,98	7,26	6,68	9,1	6,31	7,91	9,22	7,51	6,97	4,69

Mjera udaljenosti logaritama spektara dala je odlične rezultate u sva tri testa, posebice u trećem testu, gdje se u potpunosti slaže sa subjektivnim dojmovima slušatelja. Rezultati su bili vrlo slični neovisno o govornicima (uz očekivane nešto manje udaljenosti za istog govornika).

Udaljenost kepstara

Udaljenost kepstara testirana je uz kepstar signala izračunat po definiciji (tablica 3) te iz LPC koeficijenata (tablica 4). Korišteno je $L = 1{,}5p = 18$ kepstralnih koeficijenata.

U kepstar signala izračunat po definiciji rezultati se odlično slažu sa subjektivnim dojmovima slušatelja. U prvom testu udaljenosti istih glasova su male i međusobno vrlo slične. Udaljenosti $d_c(E, I)$ i $d_c(O, U)$ su manja od udaljenosti drugih parova glasova. U drugom i trećem testu rezultati su veoma slični kao i u prvom testu, uz nešto veće udaljenosti istih glasova.

Tablica 3. Udaljenosti samoglasnika korištenjem mjere udaljenosti kepstara

$d_c(S, S')$	A,A	E,E	I,I	O,O	U,U	A,E	A,I	A,O	A,U	E,I	E,O	E,U	I,O	I,U	O,U
isti govornik	1,15	**1**	1,19	1,5	1,35	6,77	7,72	4,98	5,73	4,38	6,33	6,41	6,81	6,96	4
dva muška govornika	2,94	3,27	3,54	3,08	2,75	7,28	8,27	4,78	5,55	5,56	6,73	6,95	6,84	7,09	3,9
dva govonika, muški i ženski	3,01	2,79	3,72	3,79	1,7	5,93	6,35	4,81	5,61	4,83	6,36	6,19	7,72	6,95	3,99

Korištenjem kepstra izračunatog iz LPC koeficijenata, rezultati su ostali dobri u prva dva testa, ali su se značajno pokvarili u testu s muškim i ženskim govornikom, gdje neki parovi različitih samoglasnika imaju manju udaljenost od istih samoglasnika.

Tablica 4. Udaljenosti samoglasnika korištenjem mjere udaljenosti kepstara, uz kepstar izračunat iz LPC koeficijenata

$d_{c,LPC}(S, S')$	A,A	E,E	I,I	O,O	U,U	A,E	A,I	A,O	A,U	E,I	E,O	E,U	I,O	I,U	O,U
isti govornik	1,52	1,81	1,89	2,35	1,45	13,2	15	9,7	10,9	8	11,4	10,3	12,7	12,1	6,5
dva muška govornika	5,71	5,59	5,92	5,4	3,36	12,6	14,5	11,5	11,5	9,51	13	11,2	12,8	12	6,75
dva govonika, muški i ženski	5,66	6,08	7,63	7,24	3,37	13,9	17	10,7	10,5	11,2	8,9	7,88	7,92	7,23	7,16

Težinska udaljenost kepstara

Težinska udaljenost kepstara također je testirana uz kepstar izračunat po definiciji (tablica 5) te iz LPC koeficijenata (tablica 6). Korišteno je $L = 1{,}5p = 18$ kepstralnih koeficijenata, te težinska funkcije (4.17) uz faktor $h = \frac{L}{2} = 9$.

Uz kepstar signala izračunat po definiciji, rezultati su slični kao i udaljenosti kepstara, ali uz još manje razlike u udaljenostima glasova različitih govornika. U prvom testu udaljenosti istih glasova su vrlo male i gotovo jednake za sve parove istih glasova. Udaljenosti $d_{wc}(E, I)$ i $d_{wc}(O, U)$ su manje od svih ostalih udaljenosti parova različitih glasova.

U drugom i trećem testu rezultati su veoma slični. Kao i kod udaljenosti kepstara, povećale su se udaljenosti istih glasova, dok su udaljenosti različitih glasova približno jednake kao u prvom testu, ali još uvijek dovoljno veće.

Tablica 5. Udaljenosti samoglasnika korištenjem mjere težinske udaljenosti kepstara

$d_{wc}(S, S')$	A,A	E,E	I,I	O,O	U,U	A,E	A,I	A,O	A,U	E,I	E,O	E,U	I,O	I,U	O,U
isti govornik	1,03	**1**	1,04	1,42	1,14	5,48	6,02	4,26	4,75	3,78	4,32	4,74	5,3	4,17	3,58
dva muška govornika	3,1	3,19	3,6	2,76	2,29	6,13	6,14	4	4,75	4,2	4,66	4,25	5,51	4,91	3,74
dva govonika, muški i ženski	2,84	2,92	3,93	3,13	1,49	4,83	5,05	4,47	4,78	4,18	4,32	4,69	5,53	5,31	4,12

Kao i kod udaljenosti kepstara, korištenje kepstralnih koeficijenata izračunatih iz LPC koeficijenata lagano je pokvarilo rezultate, posebice u testu s muškim i ženskim govornikom.

Tablica 6. Udaljenosti samoglasnika korištenjem mjere težinske udaljenosti kepstara, uz kepstar izračunat iz LPC koeficijenata

$d_{wc,LPC}(S, S')$	A,A	E,E	I,I	O,O	U,U	A,E	A,I	A,O	A,U	E,I	E,O	E,U	I,O	I,U	O,U
isti govornik	1,48	1,64	1,61	2,22	1,06	10,9	11,5	8,23	9,51	6,59	7,63	5,86	9,34	7,26	5,85
dva muška govornika	5,37	5,27	5,99	4,58	2,77	11,1	12,2	9,5	9,4	9,3	9	7	9,06	6,7	5,4
dva govonika, muški i ženski	4,75	5,29	6,91	5,51	3,1	10,6	11,9	8,41	8,79	9	6,7	6,6	6,3	4	6,2

Itakura-Saito mjera udaljenosti

Zbog asimetričnosti je Itakura-Saito mjera udaljenosti testirana u oba smjera, a rezultati su prikazani u tablici 7. U prvom testu kod istih glasova je udaljenost u oba smjera gotovo ista za sve snimke glasova te je stoga u tablici prikazan samo jedan rezultat. Odmah se primjećuje da Itakura-Saito mjera udaljenosti ima mnogo veći raspon između najmanje i najveće udaljenosti od svih dosad testiranih mjera. Rezultati su za većinu glasova vrlo dobri, ali problem se javlja zbog asimetričnosti – neki parovi različitih glasova, iako u jednom smjeru daju veliku udaljenost, u drugom smjeru daju mnogo manju udaljenost koja po mišljenju slušatelja nije utemeljena. Primjerice, udaljenost $d_{IS}(A, I)$ je 7 tisuća puta veća od udaljenosti $d_{IS}(I, I')$, ali je udaljenost $d_{IS}(I, A)$ samo 15,3 puta veća od $d_{IS}(I, I')$. U drugom i trećem testu, sa različitim govornicima, stanje je još lošije utoliko što postoje manje udaljenosti između nekih različitih glasova nego između istih glasova.

Međutim, zanimljivo je primjetiti da u svim slučajevima primjerice premalena udaljenost u jednom smjeru ima „protutežu“ u velikoj udaljenosti u suprotnom smjeru. Očito će simetrične mjere dobivene usrednjavanjem Itakura-Saito mjere u tim slučajevima dati mnogo bolje rezultate.

Tablica 7. Udaljenosti samoglasnika korištenjem Itakura-Saito mjere udaljenosti

$d_{IS}(S, S')$	A,A' A',A	E,E' E',E	I,I' I',I	O,O' O',O	U,U' U',U	A,E E,A	A,I I,A	A,O O,A	A,U U,A	E,I I,E	E,O O,E	E,U U,E	I,O O,I	I,U U,I	O,U U,O
isti govornik	2,5	2,08	**1**	2,55	1,69	277 41,9	7241 15,3	553 19,2	3288 16,2	279 9,6	844 20,6	1878 15,4	33,5 555	308 22,9	97 6,3
dva muška govornika	14,5 9,55	21,1 7,29	11 13,1	24,1 5,45	4,71 2,73	1532 445	8431 16,7	1313 12.5	4265 17,1	624 13,8	2140 14,5	3232 17	149 96,8	506 13,1	172 7,48
dva govonika, muški i ženski	26,9 7,7	19,9 7,3	29,6 10,6	24,1 34,1	4,12 3,95	140,7 73,7	2907 16,9	398 23,5	3810 18,8	726 9,85	1292 65,5	3964 17,1	308 1626	619 84,9	122 6,7

COSH mjera udaljenosti

Rezultati COSH mjere udaljenosti prikazani u tablici 8 zapravo su jednaki aritmetičkoj sredini udaljenosti dobivenih Itakura-Saito mjerom. Time je riješen problem velikih razlika u udaljenostima ovisno o smjeru i rezultati dobiveni COSH mjerom odlično se slažu sa dojmovima slušatelja. Velika je prednost COSH mjere veliki raspon udaljenosti – udaljenost između istih glasova barem je deset puta manja od udaljenosti različitih glasova te je stoga pogodna za upotrebu u sustavima za prepoznavanje govora.

U testovima s različitim govornicima rezultati se i dalje odlično slažu sa subjektivnim dojmom. Udaljenost $d_{COSH}(O,U)$ mnogo je veća od udaljenosti između istih glasova, ali manja od udaljenosti ostalih parova različitih glasova.

Tablica 8. Udaljenosti samoglasnika korištenjem COSH mjere udaljenosti

$d_{COSH}(S,S')$	A,A	E,E	I,I	O,O	U,U	A,E	A,I	A,O	A,U	E,I	E,O	E,U	I,O	I,U	O,U
isti govornik	2,5	2,09	**1**	2,55	1,69	160	3628	286	1672	144	432	946	294	165	51
dva muška govornika	12,1	14,2	12,1	14,9	3,72	787	4224	663	2141	319	1078	1625	122	260	89,6
dva govonika, muški i ženski	17,3	13,6	20,1	25,6	4,04	107	1462	211	1914	368	680	1491	967	352	65

Itakurina mjera udaljenosti

Rezultati testova korištenjem Itakurine mjere udaljenosti prikazani su u tablici 9. U prvom testu udaljenost istih glasova je desetak puta manja od udaljenosti različitih glasova, a udaljenosti $d_I(E,I)$ i $d_I(O,U)$ su manje od ostalih udaljenosti različitih glasova. Smjer udaljenosti nema tako velik utjecaj na udaljenost kao u slučaju Itakura-Saito mjere.

U drugom i trećem testu udaljenosti između istih samoglasnika su nekoliko puta veće nego u prvom testu, ali udaljenosti različitih glasova nisu pratile taj rast i otprilike su na istoj razini kao u prvom testu.

Tablica 9. Udaljenosti samoglasnika korištenjem Itakurine mjere udaljenosti

$d_I(S,S')$	A,A' A',A	E,E' E',E	I,I' I',I	O,O' O',O	U,U' U',U	A,E E,A	A,I I,A	A,O O,A	A,U U,A	E,I I,E	E,O O,E	E,U U,E	I,O O,I	I,U U,I	O,U U,O
isti govornik	1,05	1,44	1,81	2,85	**1**	29 24.9	35,2 22,4	21,8 27,9	22,3 23,8	15,3 12,9	25 25,9	16,7 19,6	19,3 34,1	21,4 21,6	12,6 8,6
dva muška govornika	8,15 9,27	6,11 11,7	8,72 11,5	7,89 8,91	3,38 4,07	33,5 24,6	35,1 22,5	21 19,2	21,2 21,3	18,6 19,8	23,6 23,1	18,5 18	22,2 30	22,9 17,9	13,9 8,5
dva govonika, muški i ženski	9,8 9,13	10,3 9,85	15,9 12,4	13,6 16,9	2,79 7,85	22,3 27,8	28,8 15,4	15,1 23,2	21,2 30,7	20,2 14	22,9 39,5	17,5 22,8	30,4 45,9	23,9 34,1	10,7 8,3

Mjera udaljenosti temeljena na omjeru vjerodostojnosti

Mjera udaljenosti temeljena na omjeru vjerodostojnosti rezultirala je najvećim rasponom udaljenosti sa čak 6 redova veličina između najmanjih i najvećih udaljenosti. U prvom testu

udaljenosti istih samoglasnika su vrlo male, udaljenosti $d_{LR}(E,I)$ i $d_{LR}(O,U)$ desetak puta veće, a udaljenosti ostalih parova različitih glasova još barem deset puta veće. Takvi se rezultati dobro slažu sa subjektivnim dojmom i očekivanjima.

U drugom i trećem testu, slično kao kod Itakura-Saito mjere, se pojavio problem zbog asimetričnosti. Usrednjavanjem rezultata i dobivanjem simetrične mjere kao $d_{LRsim} = \frac{d_{LR}(S,S')+d_{LR}(S',S)}{2}$ rezultati bi bili malo bolji, ali ne kao u slučaju Itakura-Saito i COSH mjera.

Tablica 10. Udaljenosti samoglasnika korištenjem mjere udaljenosti temeljene na omjeru vjerodostojnosti

$d_{LR}(S,S')$	A,A' A',A	E,E' E',E	I,I' I',I	O,O' O',O	U,U' U',U	A,E E,A	A,I I,A	A,O O,A	A,U U,A	E,I I,E	E,O O,E	E,U U,E	I,O O,I	I,U U,I	O,U U,O
isti govornik	**1**	1,54	2,32	5,49	1,15	4e4 8913	2,6e5 3626	4412 3,8e4	4421 7523	376 169	1,1e4 1,5e4	601 1926	1513 1,9e5	2971 4053	178 37
dva muška govornika	35,7 62,3	15,6 250	62 202,3	41,2 52,6	5,26 7,64	1,6e5 9384	3e5 5478	3026 1558	3534 2957	1192 2942	6090 9223	1138 919	3666 7,7e4	5118 1300	261 38,2
dva govonika, muški i ženski	61,8 74,7	86,2 61,9	462 175	266 454	3,7 51,8	4570 1284	3,7e4 360	345 7086	3129 7e4	2097 122	4655 3e6	760 9824	5,8e4 2e6	7317 2,3e5	94,3 56,1

4.5. Ocjena učinkovitosti LPC *vocodera*

Mjere udaljenosti govora koriste se i za procjenu učinkovitosti sintetiziranog govora u *vocoderima*. U ovom će se testu izmjeriti udaljenosti samoglasnika sintetiziranih LPC *vocoderom* od izvornih signala.

4.5.1. Način testiranja i rezultati

Korištenjem obrađenih mjera udaljenosti izmjerene su udaljenosti sintetiziranih glasova od izvornog signala. Glasovi su sintetizirani LPC *vocoderom* uz prediktore 6., 8., 10., 12., 14. i 16. reda. Rezultati su prikazani u tablici 11.

Tablica 11. Udaljenosti sintetiziranih samoglasnika od izvornog signala

	$A, p = 6$	$A, p = 8$	$A, p = 10$	$A, p = 12$	$A, p = 14$	$A, p = 16$
$d_2(S, S')$	0,9636	0,5654	0,4390	0,3179	0,1975	0,1520
$d_c(S, S')$	0,6188	0,5993	0,5603	0,5098	0,4975	0,4877
$d_{c,LPC}(S, S')$	0,6785	0,3727	0,2846	0,2033	0,1114	0,0653
$d_{wc}(S, S')$	3,7980	2,9978	2,7966	2,2059	2,1565	2,1483
$d_{wc,LPC}(S, S')$	8,3196	4,5010	3,6769	2,4673	1,5671	0,8651
$d_{IS}(S, S')$	0,6770	0,1925	0,1048	0,0538	0,0194	0,0115
$d_{COSH}(S, S')$	0,5725	0,1762	0,1009	0,0515	0,0197	0,0116
$d_I(S, S')$	0,5154	0,1750	0,0982	0,0513	0,0185	0,0105
$d_{LR}(S, S')$	0,6742	0,1913	0,1031	0,0527	0,0187	0,0105

Sve mjere udaljenosti dale su očekivane rezultate, gdje se udaljenost smanjuje porastom reda prediktora, ali sve manje i manje. Tako je moguće odrediti i optimalni red prediktora u *vocoderu* iznad kojeg je dobitak zanemariv, a može se optimizirati i izbor nekih drugih parametara *vocodera*.

4.6. Ocjena kvalitete kvantiziranog i kodiranog govora

Određivanje objektivne sličnosti govornih signala koristi se i za procjenu učinkovitosti u kodiranju govora. Iako se uobičajeno koristi kod *vocodera*, u ovom će se testu izmjeriti

udaljenosti izvornih signala glasova "A" i "O" od kvantiziranih signala, te od signala dobivenih adaptivnom diferencijalnom pulsno-kodnom modulacijom. Ovim će se testom vidjeti i kako se ponašaju mjere udaljenosti u slučaju malih razlika između glasova.

4.6.1. Način testiranja i rezultati

Izravna kvantizacija govornog signala provedena je skalarnom kvantizacijom s ograničenom entropijom (entropy constrained scalar quantization, *ECSQ*), uz entropiju $H(I) = 4bit$ te $H(I) = 8bit$. Za signal kodiran ADPCM-om korišteni su redovi prediktora $p = 4$ i $p = 10$, te entropije $H(I) = 4bit$ i $H(I) = 8bit$. Očekivane su veće udaljenosti signala s manjom entropijom i manjim prediktorom, te veća udaljenost izravno kvantiziranog signala od ADPCM kodiranog. Također, očekuju se i nešto veće udaljenosti u slučaju glasa "O", budući da je i omjer signala i šuma kod njega veći nego kod glasa "A". Za asimetrične mjere udaljenosti mjerene su samo udaljenosti od kvantiziranog/kodiranog glasa do izvornog, ne i u drugom smjeru.

Tablica 12. Udaljenosti izravno kvantiziranog i ADPCM kodiranog glasa "A" od izvornog signala

Glas "A"	Izravna kvantizacija		ADPCM			
	$H(I) = 4bit$	$H(I) = 8bit$	$p = 4, H = 4bit$	$p = 10, H = 4bit$	$p = 4, H = 8bit$	$p = 10, H = 8bit$
$d_2(S, S')$	0,3487	0,0016	0,0564	0,0468	0,0013	0,0009
$d_c(S, S')$	0,1595	0,0043	0,0451	0,0405	0,0024	0,0017
$d_{c,LPC}(S, S')$	0,2231	0,0153	0,0123	0,0098	0,0041	0,0033
$d_{wc}(S, S')$	1,0875	0,0288	0,3225	0,2898	0,0198	0,0115
$d_{wc,LPC}(S, S')$	0,5512	0,0383	0,0302	0,0276	0,0146	0,0083
$d_{IS}(S, S')$	0,0515	1,32e-6	0,0015	0,0011	8e-7	4e-7
$d_{COSH}(S, S')$	0,0623	1,32e-6	0,0016	0,0011	8e-7	4e-7
$d_I(S, S')$	0,1167	0	0,0025	0,0018	0	0
$d_{LR}(S, S')$	0,1238	0	0,0025	0,0018	0	0

Tablica 13. Udaljenosti izravno kvantiziranog i ADPCM kodiranog glasa "O" od izvornog signala

Glas "O"	Izravna kvantizacija		ADPCM			
	$H(I) = 4bit$	$H(I) = 8bit$	$p = 4, H = 4bit$	$p = 10, H = 4bit$	$p = 4, H = 8bit$	$p = 10, H = 8bit$
$d_2(S, S')$	1,5598	0,0356	0,3501	0,3064	0,0038	0,0007
$d_c(S, S')$	0,4672	0,0246	0,1760	0,1656	0,0057	0,0029
$d_{c,LPC}(S, S')$	0,2712	0,0253	0,2011	0,1855	0,0082	0,0067
$d_{wc}(S, S')$	3,2651	0,1829	1,2952	1,28	0,0385	0,0236
$d_{wc,LPC}(S, S')$	2,9142	0,2155	1,1985	1,1653	0,0541	0,0302
$d_{IS}(S, S')$	0,6486	6,2e-4	0,0499	0,0384	7,4e-6	2,7e-7
$d_{COSH}(S, S')$	2,1246	6,3e-4	0,0638	0,0488	7,4e-6	2,7e-7
$d_I(S, S')$	1,3589	0,0010	0,1224	0,1008	0	0
$d_{LR}(S, S')$	2,8921	0,0010	0,1302	0,1060	0	0

U tablicama 12 i 13 su prikazane udaljenosti izravno kvantiziranih te ADPCM kodiranih glasova "A" i "O" od izvornog signala. Glas "O" rezultirao je nešto većim udaljenostima od glasa "A", što se slaže i sa manjim omjerom signala i šuma u slučaju kvantiziranog ili kodiranog glasa "O" u odnosu na glas "A". Itakurina mjera udaljenosti i mjera udaljenosti temeljena na omjeru vjerodostojnosti u nekim slučajevima sa većom entropijom dale su zanemarivo male vrijednosti koje se smatraju nulom. Ostale su mjere udaljenosti dale zadovoljavajuće rezultate.

5. Zaključak

Linearno prediktivno kodiranje i kepstralna analiza pokazali su se vrlo učinkovitima u izdvajanju fonetski značajnih parametara govora. Autokorelacijska metoda najbrža je u računanju LPC koeficijenata, dok korištenje rešetkaste metode u istu svrhu nudi jednostavnu i lako modularnu blokovsku izvedbu, ali uz nešto duže vrijeme računanja.

Kepstralna analiza govora omogućila je prikaz govornog signala, nastalog konvolucijom pobudnog signala i funkcije vokalnog trakta, kao linearne kombinacije tih komponenata u kepstru. Pokazalo se i da, u svrhu mjerenja udaljenosti govornih signala, nije potrebno rekonstruirati funkciju vokalnog trakta u vremenskoj domeni (što bi značilo i nužno korištenje kompleksnog kepstra radi očuvanja informacije o fazi signala), već se uspoređuju izravno kepstralni koeficijenti realnog kepstra govornog signala.

Simetrične mjere udaljenosti (udaljenost logaritama spektara, kepstralne udaljenosti i COSH mjera udaljenosti) u testovima su dale bolje rezultate od asimetričnih.. Korištenje kepstralnih koeficijenata izračunatih iz LPC koeficijenata, iako je značajno brže od računanja kepstra po definiciji, u kepstralnim je udaljenostima dovelo do određenog pogoršanja rezultata. Itakura-Saito mjera često je rezultirala iznimno malim udaljenostima između različitih glasova u jednom smjeru (utjecaj člana V), ali velikim u drugom smjeru (utjecaj člana e^V), što je omogućilo da se usrednjavanjem dobije iznimno učinkovita COSH mjera udaljenosti. Itakurina mjera udaljenosti i mjera udaljenosti temeljena na omjeru vjerodostojnosti davale su prihvatljive rezultate, ali ipak lošije od simetričnih mjera.

6. Literatura

[1] RABINER, LAWRENCE R.; SCHAFER, RONALD W.: "Digital Processing of Speech Signals", Prentice-Hall, New Jersey, 1978.

[2] RABINER, LAWRENCE R.; BING-HWANG, JUANG: "Fundamentals of Speech Recognition", Prentice-Hall, New Jersey, 1993.

[3] DELLER JR., JOHN R.; HANSELL, JOHN H.L.; PROAKIS, JOHN G.: "Discrete-Time Processing of Speech Signals", John Wiley & sons, New York, 2000.

[4] RABINER, LAWRENCE R.; SCHAFER, RONALD W.: "Introduction to Digital Speech Processing", now Publishers Inc., Hanover, 2007.

[5] SCHNELL, KARL.: "Time-Varying Burg Method for Speech Analisys", s Interneta, http://www.eurasip.org/Proceedings/Eusipco/Eusipco2008/papers/1569102274.pdf, 6.5.2011.

[6] GULJAŠ, BORIS: "Metrički prostori", s Interneta, http://www.mathos.hr/metricki/nastavni-materijali/guljas-metprost12.pdf, 18.5.2011.

[7] UNGAR, ŠIME: "Kompleksna analiza", s Interneta, http://web.math.hr/~ungar/NASTAVA/KA/kompleksna.pdf, 12.5.2011.

[8] SENGUPTA, S.: "Linear Prediction of Speech", s Interneta, http://www.youtube.com/watch?v=4uQsp10rGKU, 2.5.2011.

[9] RABINER, LAWRENCE R.; BING-HWANG, JUANG: "Automatic Speech Recognition – A Brief History of the Technology Development", s Interneta, http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf, 12.4.2011.

Sažetak

Mjere udaljenosti u obradi govornog signala

Objektivno određivanje sličnosti dvaju govornih signala vrlo je složen problem. Kako bi se pronašla matematička udaljenost koja približno odgovara subjektivnom dojmu ljudskog uha potrebno je pronaći i usporediti samo fonetski značajne parametre govora. Korištenje spektra govornog signala u mjerenju udaljenosti daje loše rezultate jer, osim fonetski značajne prijenosne funkcije vokalnog trakta, spektar uključuje i pobudni signal koji unosi neželjenu i fonetski beznačajnu informaciju. U svrhu odvajanja prijenosne funkcije vokalnog trakta od signala pobude učinkovitima su se pokazali *all-pole* modeliranje govora pomoću linearnog prediktivnog kodiranja te kepstralna reprezentacija govora.

S obzirom na logaritamsku percepciju glasnoće zvuka kod ljudi, logično je i da se većina učinkovitih mjera udaljenosti temelji na razlikama kepstara ili logaritama spektara govornih signala. Neke od najčešće korištenih simetričnih mjera su udaljenosti logaritama spektara, kepstralna udaljenost i težinska kepstralna udaljenost. Druga porodica mjera udaljenosti temelji se na vjerodostojnosti i uključuje Itakura-Saito mjeru udaljenosti, COSH mjeru udaljenosti, Itakurinu mjeru udaljenosti te mjeru udaljenosti temeljenu na omjeru vjerodostojnosti.

Ključne riječi: govor, udaljenost, LPC, kepstar, Itakura.

Distance measures in speech signal processing

Objective speech similarity assessment is a very complex problem. In order to derive a distance function that suits the subjective judgment of sound difference, one must identify and compare only phonetically relevant speech parameters, such as vocal tract transfer function. The speech signal spectrum contains both the vocal tract function and phonetically irrelevant excitation function, and, therefore, does not provide good results. Cepstral representation and *all-pole* modeling of the speech signal have proven to be very effective in the task of extracting the vocal tract function and are regularly used in speech distance measuring.

Considering that human sound level perception is basically logarithmic, most of the effective speech distance measures are based on differences in either cepstra or logarithms of spectra. Some of the most often used symmetric distance measures are log spectral distance, cepstral distances and weighted cepstral distances. Other speech distance measures are based on likelihood and include Itakura-Saito, COSH, Itakura and likelihood ratio distance measures.

Keywords: speech, distance, LPC, cepstrum, Itakura.

Privitak A: Implementacija mjera udaljenosti u Matlabu

Udaljenost logaritama spektara

```
function udaljenost = d_log(sig1,sig2,p,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    p = 12;
end

udaljenost = 0;
for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    H1 = freqz( [1], A1(1,:), 1024);
    S1 = G1 * H1.^2;
    H2 = freqz( [1], A2(1,:), 1024);
    S2 = G1 * H2.^2;
    V = log( abs(S1)) - log( abs(S2));
    if ( isnan(trapz( ( abs(V)).^2) / 1024) ~= 1)
        udaljenost = udaljenost + sqrt( trapz( ( abs(V)).^2) / 1024);
    end
end
end
```

Udaljenost kepstara

```
function udaljenost = d_ceps(sig1,sig2,L,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    L = 18;
end
```

```
udaljenost = 0;
suma = 0;

for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    c1 = rceps( sig1sample);
    c2 = rceps( sig2sample);
    for j = 1 : L
        suma = suma + (c1(j+1)- c2(j+1))^2;
    end
    udaljenost = udaljenost + sqrt( suma);
end
end
```

Udaljenost kepstara korištenjem LPC koeficijenata

```
function udaljenost = d_cepslpc(sig1,sig2,L,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    L = 18;
end

udaljenost = 0;
suma = 0;
p = round( L*2/3);

for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    clpc1(1 : L+1) = 0;
    clpc1(1) = log(G1);
    clpc2(1:L+1) = 0;
    clpc2(1) = log(G2);
    for m = 1 : p
        clpc1(m+1) = -A1(m+1);
        clpc2(m+1) = -A2(m+1);
        for k = 1 : m-1
          clpc1(m+1) = clpc1(m+1) + (k/m) * clpc1(k+1) * (-A1(m-k+1));
          clpc2(m+1) = clpc2(m+1) + (k/m) * clpc2(k+1) * (-A2(m-k+1));
        end
    end
    for m = (p+1) : L
        for k = m-p : m-1
            clpc1(m+1) = clpc1(m+1) + (k/m) * clpc1(k+1) * (-A1(m-k+1));
```

```
            clpc2(m+1) = clpc2(m+1) + (k/m) * clpc2(k+1) * (-A2(m-k+1));
        end
    end
    for j = 1 : L
        suma = suma + (clpc1(j+1)- clpc2(j+1))^2;
    end
    udaljenost = udaljenost + sqrt( suma);
end
end
```

Težinska udaljenost kepstara

```
function udaljenost = d_wceps(sig1,sig2,L,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    L = 18;
end

udaljenost = 0;
suma = 0;

for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    c1 = rceps( sig1sample);
    c2 = rceps( sig2sample);
    for j = 1 : L
        suma = suma + ( (1+L/2*sin(j*pi/L)) * (c1(j+1)-c2(j+1)) )^2;
    end
    udaljenost = udaljenost + sqrt( suma);
end
end
```

Težinska udaljenost kepstara korištenjem LPC koeficijenata

```
function udaljenost = d_wcepslpc(sig1,sig2,L,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
```

```
    L = 18;
end

udaljenost = 0;
suma = 0;
p = round( L*2/3);

for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    clpc1(1 : L+1) = 0;
    clpc1(1) = log(G1);
    clpc2(1:L+1) = 0;
    clpc2(1) = log(G2);
    for m = 1 : p
        clpc1(m+1) = -A1(m+1);
        clpc2(m+1) = -A2(m+1);
        for k = 1 : m-1
          clpc1(m+1) = clpc1(m+1) + (k/m) * clpc1(k+1) * (-A1(m-k+1));
          clpc2(m+1) = clpc2(m+1) + (k/m) * clpc2(k+1) * (-A2(m-k+1));
        end
    end
    for m = (p+1) : L
        for k = m-p : m-1
            clpc1(m+1) = clpc1(m+1) + (k/m) * clpc1(k+1) * (-A1(m-k+1));
            clpc2(m+1) = clpc2(m+1) + (k/m) * clpc2(k+1) * (-A2(m-k+1));
        end
    end
    for j = 1 : L
        suma = suma + ( (1+L/2*sin(j*pi/L)) * (clpc1(j+1)-clpc2(j+1)) )^2;
    end
    udaljenost = udaljenost + sqrt( suma);
end
end
```

Itakura-Saito mjera udaljenosti

```
function udaljenost = d_ItakuraSaito(sig1,sig2,p,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    p = 12;
end

udaljenost = 0;
for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
```

```
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    H1 = freqz( [1], A1(1,:), 1024);
    S1 = G1 * H1.^2;
    H2 = freqz( [1], A2(1,:), 1024);
    S2 = G1 * H2.^2;
    V = log( abs(S1)) - log( abs(S2));
    udaljenost = udaljenost + trapz ( abs(S1)./abs(S2) - V - 1) / 1024;
end
end
```

COSH mjera udaljenosti

```
function udaljenost = d_COSH(sig1,sig2,p,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    p = 12;
end

udaljenost = 0;
for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    H1 = freqz( [1], A1(1,:), 1024);
    S1 = G1 * H1.^2;
    H2 = freqz( [1], A2(1,:), 1024);
    S2 = G1 * H2.^2;
    V = log( abs(S1)) - log( abs(S2));
    udaljenost = udaljenost + trapz( cosh( abs( V)) - 1) / 1024;
end
end
```

Itakurina mjera udaljenosti

```
function udaljenost = d_Itakura(sig1,sig2,p,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
```

```
    pomak = N;
end
if (nargin == 2)
    p = 12;
end

udaljenost = 0;
for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    H1 = freqz( [1], A1, 1024);
    H2 = freqz( [1], A2, 1024);
    udaljenost = udaljenost + log( trapz( (abs(H1)./abs(H2)).^2) / 1024);
end
end
```

Mjera udaljenosti temeljena na omjeru vjerodostojnosti

```
function udaljenost = d_LikelihoodRatio(sig1,sig2,p,N,pomak)

if (nargin == 4)
    pomak = N;
end
if (nargin < 4)
    N = min( length(sig1), length(sig2));
    pomak = N;
end
if (nargin == 2)
    p = 12;
end

udaljenost = 0;
for i = 1 : pomak : min( length(sig1), length(sig2)) - N + 1
    sig1sample = sig1(i : i+N-1);
    sig2sample = sig2(i : i+N-1);
    [ A1 G1 ] = lpc( sig1sample, p);
    [ A2 G2 ] = lpc( sig2sample, p);
    H1 = freqz( [1], A1(1,:), 1024);
    H2 = freqz( [1], A2(1,:), 1024);
    udaljenost = udaljenost + trapz( ( (abs(H1)./abs(H2)).^2) / 1024) - 1;
end
end
```

Privitak B: LPC *vocoder*

```
function signal_synth = lpc_vocoder(signal,p,fs,N,pomak)

if ( nargin == 4 )
    pomak = N;
end
if ( nargin < 4 )
    N = length(signal);
    pomak = N;
end
if ( nargin < 3 )
    fs = 8000;
end
if ( nargin == 1 )
    p = 12;
end

signal_synth(1 : length(signal)) = 0;
f1 = fs / 50;
f2 = fs / 500;

for i = 1 : pomak : length(signal) - N + 1
    sigsample = signal(i : i+N-1);
    r = xcorr(sigsample,f1,'coeff');
    r = r(f1+1 : 2*f1+1);
    [ rmax, index ] = max( r(f2 : f1));
    if ( rmax > 0.5 )
        FF = fs / (f2 + index - 1);
    else
        FF = 500;
    end
    if ( FF < 250)
        tff = 0 : 1/FF : 1;
        tfs = 0 : 1/fs : 1;
        pobuda = 2 * pulstran(tfs(1 : N),tff,@rectpuls,1/fs);
    else
        pobuda = 2 * ( rand(N,1) - 0.5);
    end
    [ A G ] = lpc( sigsample, p);
        signal_synth(i : i+N-1) = filter(sqrt(G),A,pobuda);
    signal_synth = ( signal_synth)';
end
```