

POREĐENJE PERFORMANSI ALGORITAMA ZA PRAĆENJE FORMANATA

Ljubomir Jovanov, Darko Pekar, *Alfanum – govorne tehnologije, Novi Sad*

Sadržaj – U ovom radu prikazan je algoritam za procenu i praćenje formantne strukture govornog signala. Opisan je algoritam za minimizaciju greške procenjenih vrednosti formantnih učestanosti, kao i algoritam za korekciju putanja formanata u vremensko-frekvencijskoj ravni. Pored toga izvršeno je kvantitativno poređenje putanja formanata dobijenih korišćenjem javno dostupnih algoritama.

1. UVOD

Parametri formantnih učestanosti daju dobar uvid u proces proizvodnja i percepcije govora i predstavljaju jedan od minimalnih skupova parametara za opis govornog signala. Struktura formantnih učestanosti predstavlja robusan skup obeležja u odnosu na šum i izobličenja u komunikacionom kanalu. Rezultati formantne analize nalaze svoju primenu u algoritmima za vizualizaciju govora, sistemima za komunikaciju u otežanim uslovima, pomagalicama za osobe oštećenog sluha i sistemima za automatsko prepoznavanje govora kao ulazni parametri koji obezbeđuju veću robusnost sistema.

Algoritam za određivanje kratkoročnih procena formantnih učestanosti zasniva se na minimizaciji greške aproksimacije. Spektar govornog signala se aproksimira određenim brojem digitalnih rezonatora koji predstavljaju rezonantne učestanosti vokalnog trakta. Zatim se algoritmom zasnovanim na dinamičkom programiranju, određuju parametri digitalnih rezonatora tako da je greška aproksimacije minimalna.

Kratkoročne procene dobijene na prethodno navedeni način ipak ne prate putanje formanata u dovoljnoj meri, pa je razvijen algoritam za korekciju putanja koji se zasniva na optimizaciji putanje u odnosu na proizvoljno odabrani skup parametara (npr. minimizacija verovatnoće «pogrešnog» prelaza, odstupanja od LPC spektra itd.). Parametri algoritma za korekciju putanje mogu se menjati u vremenu, u zavisnosti od ulaza u sistem.

2. ALGORITAM ZA ODREĐIVANJE KRATKOROČNIH PROCENA FORMANTNIH UČESTANOSTI

U ovom delu rada opisan je algoritam za određivanje kratkoročnih procena formantnih učestanosti koji se zasniva na skupu paralelnih digitalnih rezonatora.

U cilju estimacije formantnih učestanosti, vrši se podela frekventnog opsega u kome se vrši određivanje formanata na određenih broj segmenata, u okviru kojih se vrši određivanje formantnih učestanosti. U prvom koraku granice su fiksne. Za svaki od segmenata definiše se digitalni rezonator drugog reda, čija je prenosna karakteristika data Furijeovom transformacijom odgovarajućeg prediktorskog polinoma drugog reda.

$$A_k(e^{j\omega}) = 1 - \alpha_k e^{j\omega} - \beta_k e^{j2\omega} \quad (1)$$

gde su α_k i β_k realni prediktorski koeficijenti.

Moduo prenosne karakteristike rezonatora dat je sa:

$$|A_k(e^{j\omega})|^2 = 1 + \alpha_k^2 + \beta_k^2 - 2\alpha_k(1 - \beta_k)\cos\omega - 2\beta_k\cos2\omega \quad (2)$$

Prenosna funkcija rezonatora data je na slici 1.

Označimo početnu i krajnju tačku k-tog intervala sa ω_k i ω_{k-1} , respektivno. Greška predikcije rezonatora data je sa:

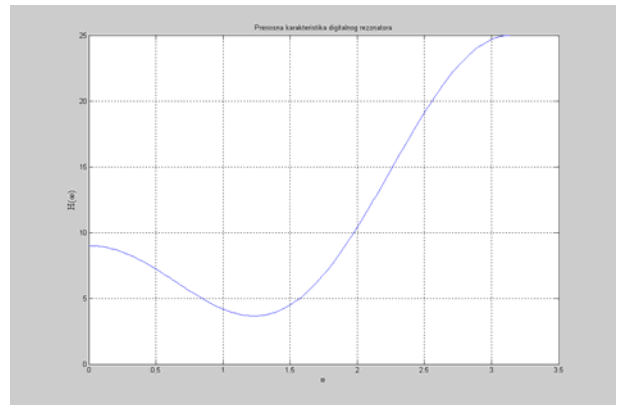
$$E(\omega_{k-1}, \omega_k | \alpha_k, \beta_k) = \frac{1}{\pi} \int_{\omega_{k-1}}^{\omega_k} |S(e^{j\omega})|^2 |A_k(e^{j\omega})|^2 d\omega \quad (3)$$

gde je $|S(e^{j\omega})|^2$ kratkotrajna spektralna gustina snage. Dakle algoritam se sastoji u pronalaženju parametara digitalnih rezonatora, tako da je zbirno potiskivanje formantnih učestanosti maksimalno.

Minimizacijom greške predikcije po α_k i β_k dobijaju se optimalne vrednosti ovih parametara:

$$\alpha_k^{opt} = \frac{r_k(0)r_k(1) - r_k(1)r_k(2)}{r_k(0)^2 - r_k(1)^2} \quad (4)$$

$$\beta_k^{opt} = \frac{r_k(0)r_k(2) - r_k(1)^2}{r_k(0)^2 - r_k(1)^2}$$



Slika 1: Prenosna funkcija digitalnog rezonatora

Ukupna greška predikcije k digitalnih rezonatora data je kao zbir grešaka pojedinih segmenata:

$$E = \sum_{k=1}^K E_{\min}(\omega_{k-1}, \omega_k) \quad (5)$$

Frekvencijski interval $[0, \pi]$ predstavljen je u svakom intervalu sa $I+1$ odbiraka, čije su frekvencije $\pi i/I$, $i=0, \dots, I$. Granice segmenata $\omega_0=0, \dots, \omega_k, \dots, \omega_K=\pi$ zamenjene su indeksima $i_0=-1, \dots, i_k, \dots, i_K=I$. Autokorelacioni koeficijenti $r_k(v)$ dati su sa:

$$r_k(v) = \frac{1}{I} \sum_{i=i_{k-1}+1}^{i_k} |S(i)|^2 \cos\left(\frac{2\pi vi}{2I}\right) \quad (6)$$

gde je:

$$S(i) = S\left[e^{j(2\pi/2I)}\right] \quad (7)$$

Cilj optimizacionog algoritma je pronalaženje granica segmenata i_1, \dots, i_{K-1} tako da je zbir:

$$\sum_{k=1}^K E_{\min}(i_{k-1} + 1, i_k)$$

minimalan.

Da bi se ubrzalo izračunavanje autokorelacionih koeficijenata datih izrazom:

$r_k(v) = T(v, i_k) - T(v, i_{k-1})$, formira se tabela sa vrednostima $T(v, i)$ na sledeći način:

$$T(v, i) = \frac{1}{I} \sum_{i'=0}^i |S(i')|^2 \cos\left(\frac{2\pi vi'}{2I}\right) \quad (8)$$

za $v=0,1,2$ i $i=0,1,\dots,I$. Uvodi se pomoćna promenljiva $F(k,i)$, koja je definisana kao greška najbolje segmentacije intervala $[0,i]$ na k delova. Podelom intervala $[0,i]$ na dva podintervala, $[0,j]$ i $[j+1,i]$, dobija se rekurentna relacija koja predstavlja osnovu optimizacionog algoritma:

$$F(k, i) = \min_j [F(k-1, j) + E_{\min}(j+1, i)] \quad (9)$$

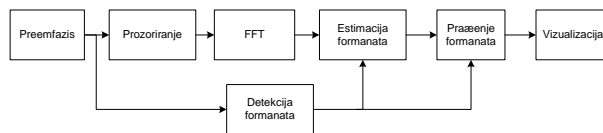
Iz izraza (9) vidi se da se najbolja segmentacija frekventnog opsega $[0,j]$ na $k-1$ intervala koristi za određivanje najbolje particije intervala $[0,i]$ na k segmenata. Da bi se omogućilo pronalaženje optimalne segmentacije, uvode se pokazivači na granice segmenata $B(k,i)$ dobijene u prethodnim koracima optimizacije.

$$B(k, i) = \arg \min_j [F(k-1, j) + E_{\min}(j+1, i)] \quad (10)$$

Rekurzivnom primenom (9) dolazi se do optimalnih granica segmenata.

Prvi korak je popunjavanje tabele sa vrednostima $T(v,i)$. Zatim se izračunavaju vrednosti $E_{\min}(j,i)$ za vrednosti $0 \leq i \leq I$ i $0 \leq j \leq i$. Nakon toga se za ceo opseg frekvencija u kom je potrebno pronaći formantne učestanosti, vrši particionisanje na $1,2,\dots,K$ segmenata i pronalaze granice koje minimizuju grešku pri svakoj podeli. Po pronalaženju optimalnih granica vrši se izračunavanje učestanosti formanata na osnovu granica α_k i β_k .

Algoritam za izdvajanje formanata predstavlja deo softverske biblioteke za digitalnu obradu signala u realnom vremenu slib. Funkcionalne celine sistema za izdvajanje formanata date su na slici 2.



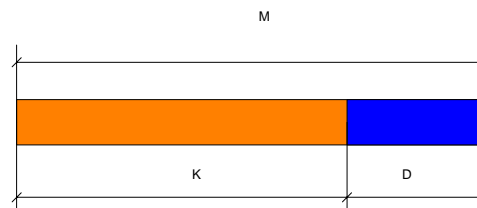
Slika 2 Sistem za izdvajanje formanata

I pored toga što se procene formantnih učestanosti u većini frejmova govornog signala slučajeva poklapaju sa vrednostima formanata određenim vizuelnim putem, algoritam ne prati kretanje formanata u vremensko-frekvencijskoj ravni na zadovoljavajući način. Da bi se postiglo bolje praćenje formanata, bilo je potrebno izvršiti korekciju putanja.

3. ALGORITAM ZA PRAĆENJE PUTANJA FORMANATA

Algoritam za praćenje formanata, na osnovu spektra i stepena zvučnosti signala vrši korekciju putanja formanata minimizacijom kriterijuma greške zadatih pri inicijalizaciji sistema.

Sama optimizacija putanja formanta u vremenu zasniva se na modifikovanom *forward-backward* algoritmu. Optimizacija putanje vrši se u realnom vremenu, po segmentima sa određenim kašnjenjem, da bi se omogućila integracija algoritma u biblioteku za obradu signala u realnom vremenu (slika 3.)



Slika 3: Način rada algoritma za optimizaciju

U prvom koraku u ulazni bafer bloka za optimizaciju upisuje se M preliminarnih procena vrednosti formanta čije se vrednosti prate. Nakon toga se vrši povratak kroz trelis i na izlaz se šalje K optimizovanih procena formanata. U sledećem intervalu u ulazni bafer se upisuje $M-D$ početnih procena i ponovo vrši optimizacija.

Optimizacija putanja vrši se posebno za svaki od k formanata. Predviđena je mogućnost optimizacije po dve vrste ulaznih parametara. U prvoj varijanti vrši se istovremeno se vrši minimizacija rastojanja procena formantnih učestanosti dobijenih na osnovu FFT signala od procena formanata dobijenih na osnovu LPC spektra i minimizacije broja pojava prelaza koji se mogu smatrati pogrešnim. Ovaj pristup daje kontinualne putanje formanata, ali ne prati dinamiku putanja u dovoljnoj meri.

Druga varijanta algoritma vrši samo minimizaciju verovatnoće «pogrešnih» prelaza. Korišćenjem ovog pristupa dobijaju se putanje koje znatno bolje prate vrednosti formanta određene vizuelnim putem. Pored toga znatno je smanjen potreban broj računarskih operacija u odnosu na prvu varijantu algoritma.

Za svaku od preliminarnih procena vrednosti formanata generiše se $2 \cdot \text{candidates_no}$ kandidata za vrednost formanata, koji se nalaze na međusobnom rastojanju $\text{candidates_distance}$. Svaki od kandidata predstavlja stanje u trelistu. Formira se matrica prelaza kojom se praktično vrši ograničavanje verovatnoće velikih varijacija vrednosti formanta. Takođe se inicijalno svakom stanju koje predstavlja preliminarnu procenu dodeljuje određena metrika.

Centralno stanje predstavlja preliminarnu procenu, dok su ostala stanja interpolirana. Metrika svakog od kandidata u trelistu dobija se na sledeći način:

$$\alpha_k(t) = \sum_{i=1}^{N_d} (\alpha_i(t-1) + p_{ik}(F(k) - F(i))) \quad (11)$$

gde su p_{ik} elementi matrice prelaza, $F(i)$ vrednost i -tog kandidata, a $\alpha_k(t)$ metrika koju je prikupio k -ti kandidat do trenutka t .

Pri svakom povratku kroz trelist polazi se od kandidata koji je akumulirao najveću metriku, i pronalaze najbolji prethodnici. Postoji mogućnost izračunavanja metrika u oba smera i računanje ukupne verovatnoće kandidata, ali se time povećava broj potrebnih izračunavanja.

Poboljšanje kvaliteta praćenja postignuto je i korišćenjem informacije o stepenu zvučnosti signala u tekućem segmentu. U slučaju zvučnog segmenta vrši se ponovna inicijalizacija algoritma za praćenje, pomoću preliminarnih procena i nastavlja sa optimizacijom.

Parametri algoritma za praćenje formanata (matrica prelaza, kriterijum po kom se vrši optimizacija, broj formanata koji se prati, broj kandidata za svaki od formanata, njihovo rastojanje, dubina optimizacije i broj odmeraka u spektru signala) zadaju se u inicijalizacionom fajlu.

4. POREĐENJE PERFORMANSI ALGORITAMA ZA PRAĆENJE FORMANATA

Tokom optimizacije, osnovni kriterijum poređenja performansi algoritama bilo je poređenje putanja formanata vizuelnim putem. Pri izboru optimalnih parametara algoritama za praćenje, prvenstveno su posmatrana dva kriterijuma: koliko verno algoritmi prate putanje formanata, s obzirom na spektar signala i na izgovorenu sekvencu i kontinualnost putanja. Ipak u cilju kvantitativnog poređenja dobijenih rezultata definisane su dve mere odstupanja putanja formanata:

- srednja kvadratna greška:

$$E_{\text{MSE}} = \sqrt{\frac{1}{N-1} \sum_{n=0}^{N-1} (\tilde{F}_{\text{Cn}} - \tilde{F}_{\text{Sn}})^2} \quad (12)$$

- srednja apsolutna greška:

$$E_{\text{MAE}} = \frac{1}{N} \sum_{n=0}^{N-1} |\tilde{F}_{\text{Cn}} - \tilde{F}_{\text{Sn}}|, \quad (13)$$

gde je N -broj frejmova u signalu, \tilde{F}_{C} – vektor izračunatih učestanosti formanata i \tilde{F}_{S} -vektor ručno labeliranih učestanosti formanata. Performanse algoritama poređene su

na govornoj sekvenci »la«, na kojoj su ručno obeleženi formanti u za to posebno implementiranom softveru. Izvršeno je poređenje performansi sledećih algoritama:

- algoritam za estimaciju učestanosti formanata na osnovu maksimuma LPC spektra
- algoritam za estimaciju učestanosti formanata na osnovu korena LPC spektra
- algoritam Rabinera i Šefera
- algoritam implementiran u programskom paketu Praat (Burg)
- predloženi algoritam

Spektar signala, ručno labelirane putanje formanata, kao i putanje dobijene korišćenjem predloženog algoritma i algoritma implementiranog u programskom paketu Praat prikazani su na slici 5. Vrednosti srednje apsolutne i srednje kvadratne greške date su u tabeli 1.

	1. maksimumi LPC spektra	2. koreni LPC spektra	3. Rabiner	4. Praat	5. DR.
MAE	56.82	83.73	137.0	53.9	31
MSE	109.56	171.31	274.2	93.57	71.9

Tabela 1: Srednja apsolutna i srednja kvadratna greška

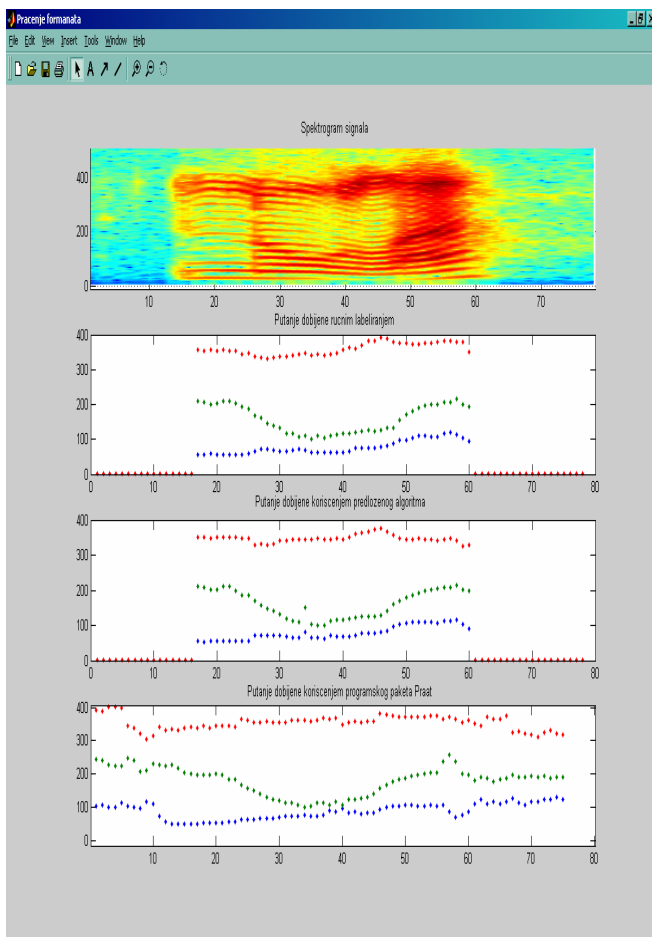
Putanje formanata dobijene korišćenjem algoritama (1-3) postprocesirane su korišćenjem algoritma opisanog u referenci [4], a Praat-ov algoritam koristi modifikovani Viterbijev algoritam. Modifikovani Viterbijev algoritam implementiran u programskom paketu Praat formira metrike na osnovu odstupanja dobijenih procena od referentnih učestanosti za dati formant, razlika dobijenih vrednosti formanata u susednim frejmovima i širine dobijenih procena formanata.

Sekvenca na kojoj su poređene performanse ima veliku dinamiku putanja formanata. Najveće odstupanje u odnosu na ručno labelirane putanje imaju algoritmi 2 i 3, što je i konstatovano u literaturi.

Algoritmi 1 i 4 daju veoma slične rezultate, što je i bilo očekivano, s obzirom da se značajnije razlikuju samo u algoritmu za praćenje putanja. Najmanje odstupanje dobijeno je korišćenjem predloženog algoritma. Slični odnosi performansi opaženi su i vizuelnim putem na većem broju sekvenci, s tim da su putanje dobijene pomoću Praat softvera i predloženim algoritmom sličnije nego što se može zaključiti na osnovu grešaka datih u tabeli 1.

3. ZAKLJUČAK

Algoritam za praćenje učestanosti formanata opisan u ovom radu predstavlja kombinaciju pristupa koji su do sada dali najbolje rezultate [2] u pogledu početne procene učestanosti formanata i algoritama za estimaciju nepoznate sekvence. Osnovna prednost algoritma za dobijanje početne procene u odnosu na druge slične algoritme je to što nije potrebna početna obuka modela za procenu učestanosti formanata [3] i relativno visoka pouzdanost. Naknadna obrada putanja u velikoj meri doprinosi kvalitetu algoritma, jer su dobijene putanje koje vernije prate formantne učestanosti.



Slika 4: Spektrogram signala, ručno labelirane putanje formana, putanje dobijene predlozenim algoritmom i programskim paketom Praat

LITERATURA

- [1] R. Obradović, D. Pekar: C++ Library for Signal Processing – SLIB, DOGS, Novi Sad, 2000.
- [2] Lutz Welling and Hermann: Ney, Formant Estimation for Speech Recognition, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 6, NO. 1, JANUARY 1998
- [3] Gary E. Kopec: Formant Tracking Using HMM and Vector Quantization, IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. ASSP-34, NO. 4, AUGUST 1986
- [4] Phillip Schmid: Explicit N-best Formant Features For Segmen-Based Speech Recognition, PhD Thesis 1996.
- [5] <http://www.praat.org>

Abstract – This paper presents the algorithm for formant tracking based on dynamic programming and Viterbi search. The paper describes the algorithm for formant frequencies calculation in base frame of speech signal, optimum sequence search algorithm and gives an overview of various formant tracking algorithms performance.

FORMANT TRACKING ALGORITHMS PERFORMANCE EVALUATION

Ljubomir Jovanov, Darko Pekar