

N. Bijedić, D. Lalović, A. Vlahov,
I. Zeljković, V. Zlatarov
Laboratorija za teorijsku fiziku, Institut "B. Kidrič"
Vinča - Beograd

REFERAT

R. Radošević
VP 2237 Beograd

ANALIZA SEGMENTACIJE GOVORNOG SIGNALA NA
PRELAZNE I STACIONARNE OBLASTI FONEMA

AN ANALYSIS OF SPEECH SIGNAL SEGMENTATION
ON TRANSIENT AND STATIONARY REGIONS OF THE
PHONEMS

SADRŽAJ - U radu su dati rezultati analize segmentacije govornog signala na prelazne i stacionarne oblasti fonema. Uzcrak govora sastojao se od 125 izolovanih reči koje je izgovaralo 7 govornika. Izvršena je preliminarna analiza prepoznavanja fonema na osnovu formantnih parametara i raspodele snage signala po specifičnim opsezima frekvencija.

ABSTRACT - The results of the analysis of the segmentation of the speech signal in the transition and stationary regions of phonems are given. The speech sample consisted of the 125 isolated words spoken by seven speakers. We have done the preliminary analyzes of the phonem recognition on the basis of formant parameters and power distribution in specific frequency bands.

1. UVOD

Neprekidni govor se može posmatrati kao niz fonema iz konačnog skupa fonema za određeni jezik. Sa akustičke tačke gledišta, svakom fonemu odgovara skup akustičkih oblika govornog signala. Fonemi se mogu razlikovati međjusobno po akustičkim osobinama svojstvenim skupovima akustičkih oblika na osnovu kojih se ti skupovi međjusobno razdvajaju. Za razne jezike postoji različita relacija između fonema i fonetskih simbola, odnosno postoji različit broj kako fonema tako i fonetskih simbola. Srpsko-hrvatski jezik je fonetski, ima 30 fonema i isto toliko fonetskih simbola.

Prepoznavanje fonema se vrši na osnovu informacionih karakteristika svakog fonema po kojima se on razlikuje od ostalih. To

znači, treba odrediti skup parametara i njihove vrednosti koje je moguće izmeriti u govornom signalu i pomoću kojih se može jednoznačno odrediti da određena oblast (segment) signala odgovara određenom fonemu. Da bi se signal mogao izučavati i obradivati u ovom smislu potrebno je prethodno odrediti oblast u signalu koja odgovara nekom od fonema. Naime, poznato je [1] da se kontinuirani govor proizvodi promenom konfiguracije sistema za proizvodjenje govora na osnovu instrukcija iz centralnog nervnog sistema. Sistem za proizvodjenje govora se kontinuirano menja. Zbog toga ne postoji diskontinuitet u trajektoriji informacionih parametara na prelazu između fonema. Prema tome, u govoru imamo stacionarne oblasti koje odgovaraju i karakterišu foneme i prelazne oblasti koje karakterišu prelaz iz prethodnog fonema u sledeći. U stacionarnim oblastima je promena informacionih parametara relativno mala. Međutim, zbog inercije (nepokretnosti) sistema za proizvodjenje govora prelazne oblasti su dosta duge. Ako je fonem u izgovoru kratkotrajan (što je dosta česta pojava) nedovoljna je dužina stacionarne oblasti signala tako da je otežena identifikacija fonema. U tom slučaju je nužno da se izučavaju i informacioni parametri u prelaznim oblastima da bi se dobili dodatni podaci za identifikaciju fonema.

Iz navedenih razloga, prepoznavanje fonema zahteva prethodnu segmentaciju govornog signala na prelazne i stacionarne oblasti a zatim na oblasti koje odgovaraju fonemu. Granica između dva fonema, koja se nalazi u prelaznim oblastima, najčešće se nemože tačno odrediti. Zavisno od potreba određuju se kriterijumi za njeno određivanje. Najčešći kriterijum je najbrža promena informacionih parametara. U nekim slučajevima za dati fonem postoje pored prelaznih i stacionarne oblasti i oblast u kojoj su promene parametara relativno velike i česte i njih smo nazvali nestacionarne.

Cilj ovoga rada je da se utvrde vrednosti informacionih parametara na osnovu kojih se može prepoznati prelaz između dva određena fonema ili grupe fonema, i kriterijumi za prepoznavanje stacionarnih (i nestacionarnih) oblasti.

Izučavan je uzorak govora od 125 slogova i reči izgovorenih od 7 govornika (5 muških i 2 ženska). Govorni signal je filtriran

na 5 KHz i digitalizovan odmeravanjem sa 10 KHz i kvantizacijom 8 bita po uzorku. Automatskom obradom uzorka na računaru primenom modela linearne predikcije [5] određene su vrednosti i trajektorije informacionih parametara. Vizuelnom analizom ovih rezultata definisane su mogućnosti prepoznavanja prelaznih oblasti između određenih grupa fonema i kriterijumi za utvrđivanje stacionarnih oblasti.

2. IZBOR INFORMACIONIH PARAMETARA, IZBOR PRAGOVA I PODELA FONEMA NA GRUPE

Izučavanjem probnog uzorka govora od 25 i rezultata iz [2] slogova izabrani su sledeći parametri.

1. Ukupna snaga signala E_1 u opsegu 0-5 KHz.
2. Razlika snage E_1 i snage E_2 u području od 0.64-2.8 KHz nazvana Δ_2 .
3. Odnos snaga u opsezima 0-0.9 KHz i 3.5-5 KHz nazvan p.
4. Formantne učestanosti prva tri formanta, F_1 , F_2 i F_3 .

Parametri snaga su računati na svakih 5 ms a formantne učestanosti na svakih 10 ms. Treba zapaziti da informativnost snaga zavisi od osnovne učestanosti, tj. u kom delu pič periode se ram koji se obradjuje nalazi. Ukoliko je ram približno jednak ili veći od pič periode, ovaj uticaj se može zanemariti.

Fonemi su podeljeni u sledećih 7 grupa: V - vokali (A,E, I,O,U), L - likvidi (J,L,LJ,R,V), N - nazali (M,N,NJ), A - afrikati (C,Č,Ć,DJ,DŽ), F_2 - zvučni frikativi (Z,Ž), F_D - bezvučni frikativi (F,H,S,Š) i P - plozivi (B,P,D,T,G,K).

Opšti kriterijum za prepoznavanje prelazne oblasti je da bar jedan od parametara ima promenu iznad zadatog praga i ako u okviru fonema ne postoji slična promena toga parametra. Prethodna analiza je pokazala da je to teško obezbediti za likvide i zvučne frikative. Zato smo uveli istovremene kriterijume za prepoznavanje i prelaznih i stacionarnih oblasti. Ustanovljeni su pragovi: $|\Delta F_1|=100$ Hz, $|\Delta F_2|=200$ Hz, $|\Delta F_3|=500$ Hz, $|\Delta E_1|=8$ dB, $|\Delta \Delta_2|=5$ dB i $|\Delta p|=10$ dB.

Ukoliko su promene iznad ovih pragova prepoznaje se prelazna oblast a ukoliko su ispod prepoznaje se stacionarna oblast.

3. REZULTATI

Na trajektorijama parametara dobijenih automatskom obradom vizuelno je ustanovljeno 2628 prelaza poznavajući redosled fonema

i pauze medju rečima. Uvedena promena parametara je ustanovljena u 2617 prelaza a ta promena ni jednog od parametara nije ustanovljena u sledećim slučajevima prelaza: $V \rightarrow L$ 3 prelaza od ukupno 217, $V \rightarrow N$ 1 od ukupno 98, $P \rightarrow N$ 1 od 7, $F_D \rightarrow A$ 2 od 7, $P \rightarrow A$ 2 od 14, $F_B \rightarrow F_B$ 1 od 7, $P \rightarrow F_B$ 1 od 7. Najbrojnije uvedene promene na prelazima imaju parametri E_1 (2079), F_2 (1792) i p (1764). Zbog toga su oni najznačajniji u segmentaciji.

Rezultati analize dati su u tabeli 1. U polja tabele upisani su brojevi slučajeva kada je premašen prag za odgovarajući parametar. U zadnjoj koloni dati su ukupni brojevi slučajeva kada svaki od pragova u prelaznoj oblasti nije premašen. Podaci su dati za jednog govornika i ukupno za svih 7 govornika.

Prepoznavanje fonema

U ovoj analizi nije izdvajana osnovna frekvencija, ali se zvučnost segmenta može odrediti na osnovu postojanja istaknutog formanta u opsegu do 1 KHz i parametra p [4]. Prag za parametar p , koji bi određivao zvučnost fonema jako zavisi od govornika. Globalno, može se konstatovati da zvučni fonemi, izuzev I, J, Z, Ž, Dj i DŽ, imaju $p > 0$ dok bezvučni fonemi imaju $p < 0$. Za pojedinačnog govornika mogu se postaviti tačnije granice.

Vokali u reči imaju dominantnu snagu, platoi u dijagramu $E_1(t)$, stabilne i konstantne formantne trajektorije. Međutim, trajektorija drugog formanta kod fonema O i U često nestaje, stapa se sa trajektorijom prvog formanta. Radi poredjenja u tabeli 2. date su srednje vrednosti položaja formantata za vokale u 125 reči izgovorena od 7 govornika (uzorak 1) i vokala izgovorenih izolovano od 30 govornika (uzorak 2).

Bitna razlika u položajima formantata za ova dva uzorka postoji samo u slučaju drugog formanta fonema A i U, što se može smatrati kao indikacija uticaja okoline na položaj F_2 kod pomenutih fonema. Međutim, opseg promene položaja formantata vokala u reči za razne govornike je znatan tako da je skoro onemogućeno međusobno razlikovanje vokala nezavisno od govornika. Za pojedinačnog govornika vokali se mogu međusobno razlikovati po položaju prva tri formanta.

Vrsta prelaza	Broj govornika	Broj prelaza	$\Delta E1 \geq 8$	$\Delta \Delta 2 \geq 5$	$\Delta P \geq 10$	$\Delta F1 \geq 100$	$\Delta F2 \geq 200$	$\Delta F3 \geq 500$	Nije identifikovan prelaz
V→V	1	5	1	2	1	1	5	1	
	7	37	1	10	1	20	30	11	
V→L	1	21	21	4	12	10	17	5	
	7	217	145	24	125	120	124	71	1
V→N	1	14	5	2	4	2	10	4	
	7	68	62	20	63	66	77	46	1
V→A	1	24	20	0	10	22	15	10	
	7	147	107	10	117	115	105	106	
V→F _Z	1	15	10	6	11	7	8	5	
	7	90	70	45	74	50	54	30	
V→F _B	1	22	20	7	22	18	14	12	
	7	152	122	64	122	122	125	50	
F→V	1	22	22	17	17	22	20	17	
	7	221	201	71	150	164	154	125	
L→L	1	1	1	0	0	0	1	0	
	7	21	20	1	5	15	5	5	
L→N	1	2	0	0	0	0	1	1	
	7	14	1	1	5	6	4	7	
L→A	1	7	7	1	2	2	7	4	
	7	46	23	10	21	22	18	16	
L→F _B	1	1	1	4	1	1	1	1	
	7	18	10	22	1	24	14	15	
L→P	1	6	5	2	1	3	2	4	
	7	42	25	15	21	25	17	22	
V→P	1	21	20	1	14	20	15	20	
	7	216	194	27	145	141	133	117	
L→V	1	44	33	11	14	15	25	12	
	7	306	182	20	154	107	225	118	
N→V	1	10	10	5	5	14	10	7	
	7	127	90	41	82	100	122	51	
A→V	1	22	22	22	18	22	24	11	
	7	202	154	150	162	186	167	104	
F _Z →V	1	6	1	4	1	1	2	1	
	7	42	22	21	32	30	25	14	
F _B →V	1	22	10	7	10	10	6	7	
	7	84	62	37	77	77	70	46	
N→L	1	1	0	0	1	0	1	1	
	7	7	5	1	5	6	4	4	
A→L	1	1	1	1	1	1	2	1	
	7	21	10	5	17	20	18	6	
F _Z →L	1	2	2	1	2	2	2	1	
	7	14	11	4	11	10	6	7	
F _B →L	1	5	7	1	7	5	7	4	
	7	26	50	25	52	52	20	20	
P→L	1	18	14	3	10	12	11	5	
	7	124	101	27	101	105	105	20	
N→A	1	1	1	1	1	1	0	0	
	7	7	7	2	6	4	4	3	
N→F _B	1	1	0	0	0	0	1	1	
	7	7	5	5	5	3	5	5	
N→P	1	2	0	0	0	1	2	1	
	7	14	12	2	12	11	11	7	
A→N	1	2	2	2	2	2	1	0	
	7	14	11	6	13	14	13	11	
F _Z →N	1	1	1	1	1	1	0	0	
	7	7	5	6	6	1	4	4	
F _B →N	1	3	3	1	3	3	2	1	
	7	21	19	11	20	20	19	13	
P→N	1	1	0	1	1	1	1	0	
	7	7	5	3	6	4	4	2	1
A→P	1	3	3	0	2	1	0	0	
	7	21	20	6	14	9	7	6	
F _B →A	1	1	0	0	0	0	0	0	1
	7	7	5	1	0	1	2	2	2
P→A	1	2	1	1	1	0	0	0	
	7	14	11	6	6	4	4	4	2
F _Z →P	1	4	4	0	1	3	2	0	
	7	28	27	13	21	17	20	13	
F _B →F _B	1	1	0	0	0	0	0	0	1
	7	7	5	0	1	2	1	1	1
F _B →P	1	12	9	6	4	2	0	0	
	7	64	60	32	21	6	6	2	
P→F _B	1	1	1	1	0	0	0	0	
	7	3	2	4	2	5	2	0	1
F→P	1	1	1	0	1	1	1	1	
	7	21	14	5	13	11	11	1	

Tabela 2.

Vo- kal	u z o r a k 1							u z o r a k 2		
	N	F ₁	F ₂	F ₃	V ₁	V ₂	V ₃	F ₁	F ₂	F ₃
A	50	764	1445	2440	14	18	7	767	1199	2338
E	41	498	1870	2558	14	16	10	422	1958	2525
I	20	322	2241	2967	18	14	13	265	2210	2928
O	30	553	974	2405	21	18	6	483	802	2491
U	27	367	851	2418	23	15	6	345	672	2320

N - broj pojavljivanja datog fonema u uzorku 1

F₁, V₁ - položaji formanta u Hz, amplitude u dB

i = (1,2,3)

Likvidi se mogu prepoznati po neprekidnim formantnim trajektorijama koje se glatko spajaju sa formantnim trajektorijama okolnih fonema. E₁(t) ima doline ili padinu. Fonem R, koji u srpsko-hrvatskom jeziku može da ima i ulogu vokala, prepoznaje se po formantima F₁-500; F₂-1500 i F₃-2400. E₁(t) ima oštar minimum širine 20 do 40 ms, dubine i do 15 dB, kod svih govornika i bez obzira na mesto fonema R u reči.

Nazali se prepoznaju po istaknutom formantu F₁-250 Hz i F₂-1200 Hz a E₁(t) ima dolinu ili padinu. Položaji ovih formanta slabo zavise od okoline.

Frikativi zvučni imaju nivo snage kao vokali ali se od ovih razlikuju po niskom nivou parametra p. Položaji formanta jako zavise od okoline.

Frikativi bezvučni mogu se podeliti po parametru E₁ u glasne (S,Š) i manje glasne (F,H). Imaju nizak nivo parametra p ali nemaju formante.

Plozivi se izdvajaju po karakterističnoj pauzi u eksploziji. Zvučni plozivi (B,D,G) imaju "formant" na približno 200 Hz po čemu se mogu razlikovati od bezvučnih. Kod ploziva na kraju reči ova se razlika obično gubi.

Afrikati se izdvajaju po karakterističnoj pauzi koja ima viši nivo po E₁ nego kod ploziva - kvazipauza. Iza kvazipauze

sledi deo sa povišenim nivoom snage koji traje više od 50 ms - kvazi eksplozija. U prepoznavanju afrikata bitnu ulogu igra parametar Δ_2 , zvučni afrikati DJ i DŽ za razliku od bezvučnih imaju u kvazi-eksploziji formante.

4. ZAKLJUČAK

U ovom istraživanju delimično korišćeni su parametri opisani u [4]. Rezultati koje smo dobili za srpsko-hrvatski jezik, slični su onima za engleski [4], u slučajevima fonema koji se mogu upoređivati. To se pre svega odnosi na karakteristike: zvučno - bezvučno, nazalnost, frikacija, pauza i eksplozija kao i na oblik promene snage signala u toku izgovaranja reči.

U daljem istraživanju, dobijeni rezultati će korisno poslužiti pri formulisanju algoritama za automatsku segmentaciju reči na foneme i prepoznavanje fonema.

REFERENCE:

1. J.L. Flanagan: SPEECH ANALYSIS, SYNTHESIS AND PERCEPTION, Springer - Verlag 1972
2. F. Jelinek: CONTINUOUS SPEECH RECOGNITION BY STATISTICAL METHODS, Proc. of the IEE, Vol. 64 (1976) p. 532
3. R. Schwartz, J. Makhoul: WHERE THE PHONEMES ARE: DEALING WITH AMBIGUITY ACOUSTIC-PHONETIC RECOGNITION, IEEE Trans. on Acous., Speech, and Sign. Proces., Vol. ASSP-23, No 1,
4. C.J. Weinstein: A SYSTEM FOR ACOUSTIC-PHONETIC ANALYSIS OF CONTINUOUS SPEECH, IEEE Trans. on Acoust., Speech and Signal Proces., Vol. ASSP 23, No 1 (1975) p. 54
5. J.D. Markel, A.H. Gray, Jr: LINEAR PREDICTION OF SPEECH (1976)

U ovim istraživanjima učestvovali su i Koković Milojko, Stojilović Dragomir i Jovanović Goran.