

UDC 612.78:004.934  
DOI 10.18485/primling.2015.16.1

**Slobodan Jovičić**

Centar za unapređenje životnih aktivnosti, Beograd  
Elektrotehnički fakultet, Univerzitet u Beogradu

**Zorka Kašić**

Fakultet za specijalnu edukaciju i rehabilitaciju  
Univerzitet u Beogradu

## SEGMENTACIJA GLASOVA U GOVORNOM IZRAZU: PROBLEMI I PRINCIPI<sup>1</sup>

**Sažetak:** U radu se razmatra pitanje segmentacije glasova u kontinuiranom govoru. Date su preporuke za segmentaciju određenih klasifikacionih grupa glasova u međusobnom kontaktu. Kao najznačajniji problemi u ovom postupku navedeni su bezvučni plozivi u inicijalnoj poziciji, koartikulacioni efekti u kontaktu vokala sa konsonantima i sonantima, kao i problem obezvučavanje glasova u finalnoj poziciji. Preporuke mogu poslužiti kao osnova za dalja istraživanja iznetih gledišta i kao podsticaj za razvoj matematičkih algoritama prilikom primene informacionih tehnologija u govornoj komunikaciji. Na kraju rada ukazuje se na činjenicu da pitanje segmentacije treba razmatrati ne samo u vremenskoj dimenziji već i u trodimenzionalnom prostoru vreme-spektar-intenzitet.

**Ključne reči:** segmentacija glasova, principi segmentacije, koartikulacija, trodimenzionalna segmentacija, klasifikacione grupe glasova.

### UVOD

U govornom toku, odnosno kontinuiranom prirodnom govornom izrazu, glasovi se skoro nikada ne artikulišu pojedinačno. Kao najmanje izgovorne jedinice koje se mogu analizom izdvojiti iz reči nekog jezika, odnosno najmanji primetni diskretni segmenti zvuka u kontinuiranom govornom toku (Simeon, 1969: I, 413), glasovi se grupišu u veće celine u okviru kojih se produkuju. Veća govorna celina od glasa je slog, koji, najjednostavnije rečeno, predstavlja jedinstvenu artikulacionu celinu. Sledeća u fonetskoj hijerarhiji je izgovorna (fonetska) reč, koju čine svi segmenti u okviru jedne akcenatske celine.

---

<sup>1</sup> Rad je realizovan u okviru projekata *Interdisciplinarna istraživanja kvaliteta verbalne komunikacije* OI 178027 i *Jezici i kulture u vremenu i prostoru* OI 178002 koje finansira Ministarstvo prosvete, nauke i tehnološkog razvoja Republike Srbije.

Ona može da se poklapa sa leksemom, ali u sebi može da sadrži i dve ili više leksema od kojih je samo jedna akcentovana (kao u primeru: *smejao sam mu se*). Najveća izgovorna (fonetska) celina je fonetska fraza - nju predstavlja deo govora omeđen dvema disajnim pauzama (Kašić, 2000).

Grupisani u pomenute govorne segmente, glasovi međusobno utiču jedni na druge, što dovodi do različitih koartikulacionih pojava, odnosno glasovnih promena. Promene glasova, koje podrazumevaju veći koartikulacioni efekat tokom kojeg glas prelazi u artikulaciono polje drugog glasa, u okviru reči (lekseme) uglavnom su stalne, pa se one u normiranim jezicima sa fonološkim pravopisom beleže i u pismu. Eksperimentalna istraživanja pokazala su da je navedeni tip promena stalan i u okviru izgovorne (fonetske) reči. Konstatovano je da kod dobrih govornika nema pauze unutar izgovorne reči koju čine proklitika, enklitika i akcentogena reč. Pošto se u govoru ne javlja pauza, obavezno dolazi do različitog međusobnog uticaja glasova u dodiru, što uslovljava promene. Te promene mogu biti samo pomeranja unutar artikulacionog polja istoga glasa, a mogu se javiti i kvalitativno novi glasovi, može doći do gubljenja glasa ili sažimanja u jedan glas produženog trajanja (Kašić, 1980, 1985), ali se ove promene i u jeziku sa fonološkim pravopisom ne beleže u pismu. S obzirom na činjenicu da organizovanost govora zavisi od više činilaca, grupisanje glasova u fonetske jedinice, hijerarhijski više od sloga, nije uvek identično, čak i kod istog govornika. To dalje znači da dodiri glasova s kraja jedne i početka druge lekseme mogu biti fakultativni, te tako fakultativne mogu biti i promene glasova uslovljene ovim dodirom (Kašić, 1990). Pomeranja glasa unutar artikulaciono-akustičkog polja su automatizovana u okviru artikulacione baze, a nemaju skoro nikakav značaj u percepciji govora. Netranirani slušaoci u eksperimentima te promene uopšte nisu registrovali, iako su spektrografskom analizom utvrđene (Kašić, 1997).

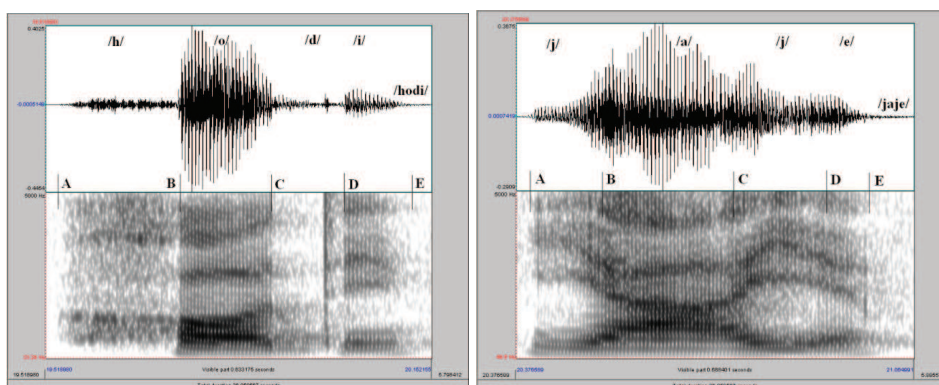
U kojoj meri će ove promene biti zastupljene u nekom jeziku zavisi od njegove artikulacione baze, odnosno od sistema automatizovanih artikulacionih navika izvornih govornika toga jezika. U svakom slučaju, glasovi su u većim govornim segmentima međusobno povezani, odnosno u govornom signalu spojeni, a sagovornici ih percipiraju kao razlikovne jedinice u sklopu razumevanja poruke (Shriberg et al., 2008).

Međutim, pored ove subjektivne segmentacije svojstvene svakom čoveku postoji potreba u mnogim naučnim disciplinama - od fonetike do savremenih govornih tehnologija, a pogotovo u ovim poslednjim, da se segmentacija i objektivno vrši na samom govornom signalu (Huang et al., 2001, Bilibajkić et al., 2011). Objektivna segmentacija podrazumeva računarsku (softversku) analizu govornog signala i primenu određenih algoritama u njegovoj obradi, koji prema zadatim kriterijumima treba da obave ovaj zadatak (Benesty et al., 2008). I sada se dolazi do suštine pitanja segmentacije glasova: kako formulisati kriterijume na osnovu kojih se mogu realizovati pouzdani algoritmi segmentacije? Ali, ovo je samo osnovno pitanje iza koga sledi čitav niz drugih kao što su: kako na segmentaciju utiču različiti glasovni spojevi; kako kod istog para glasova na segmentaciju utiču lingvistički elementi kao što su kontekst, slogovni akcenat, koartikulacija, prozodija; kakav je uticaj psihoemotivnih varijacija govornika u govornom činu; itd, itd. Kao primer u odgovoru na prvo pitanje na slici 1 prikazani su snimci dve reči "hodi" i "jaje"; prikazani su talasna forma govornog signala i ispod njega spektrogram.

U reči "hodi" artikulacioni spojevi susednih glasova su, u oba domena amplitudskom i spektralnom, veoma dobro izdiferencirani; ostaje pitanje definicije početka i kraja

ove reči. Drugi primer, reč “jaje”, znatno je komplikovaniji jer se javljaju spojevi zvučnih glasova sonanta /j/ i vokala /a/ i /e/. U amplitudskom domenu je skoro nemoguće odrediti granice susednih glasova, dok spektrogram pokazuje razlog navedenom stavu, a to je jaka koartikulacija vidljiva u tranzicijama formanata u širem vremenskom segmentu.

Slični primeri za svako postavljeno pitanje, i za mnoga druga pitanja, samo potvrđuju činjenicu da je segmentacija glasova, davnašnja istraživačka tema, tek delimično rešena pa samim tim i dalje zanimljiva i značajna u navedenim naučnim disciplinama (Mermelstein, 1975; Glass and Zue, 1988; Hatazaki et al., 1989). Ovaj rad je posvećen problemima segmentacije glasova i polazi od definicije problema segmentacije kao preuslova za postavljanje opštih načela za segmentaciju određenih klasifikacionih grupa glasova u međusobnim spojevima. Sledi više primera primene specificiranih načela i na kraju podsticajna diskusija o segmentaciji nestandardnog govora kao i potrebi multidimenzionalnog sagledavanja problema segmentacije.



Slika 1. Talasne forme i spektrogrami reči “hodi” i “jaje”.

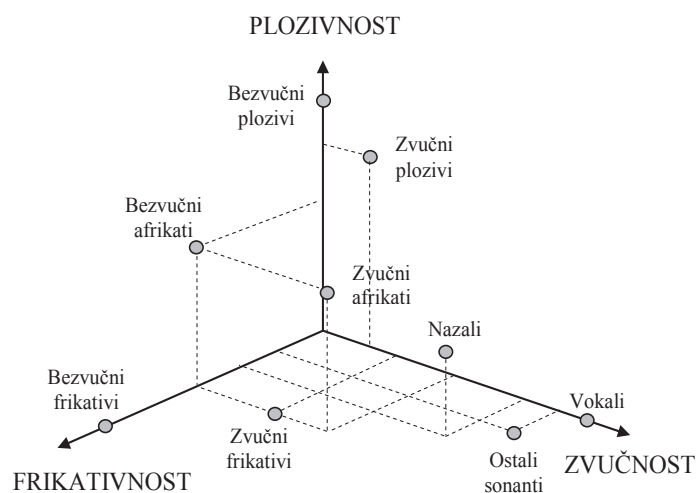
## DEFINICIJA PROBLEMA SEGMENTACIJE

Primeri na slici 1 nameću logično pitanje: kako vršiti segmentaciju sekvencijalnog niza glasova? Prirodan odgovor, kako je i naznačeno u snimcima na slici 1, jeste da segmentaciju treba vršiti u vremenskom domenu, upravo onako kako čovekov govorni aparat generiše glasove. To znači da treba postaviti markere na vremenskoj osi koji razdvajaju govorne jedinice, u ovom slučaju glasove. Koliko god ovakav pristup zadatku segmentacije izgleda jednostavan (prirodan) on nosi sa sobom niz objektivnih problema. Tri osnovna momenta čine problem segmentacije veoma kompleksnim.

- (i) Govorni segmenti su veoma podložni varijabilnostima produkcione i individualne prirode. Nikada govornik jedan govorni događaj ne može ponoviti na identičan način, a i kad ga ponovi, ponovljena verzija veoma zavisi od paralingvističkih fenomena kao što su psihoemotivno stanje govornika, fiziološke varijacije kvaliteta glasa, a i samo ponavljanje predstavlja novu komunikativnu situaciju.

- (ii) Prožimanje dva susedna glasa je neminovno, tzv. koartikulacioni fenomen, i nastaje kao posledica inercije artikulatora, te se postavljaju pitanja kao što su: kada nastaje/nestaje koartikulacioni segment imajući u vidu i same promene unutar glasova; gde je i da li postoji granica u kontaktu između dva susedna glasa; da li pojedini deo koartikulacionog segmenta pripada više ili manje jednom ili drugom glasu i šta bi takva ocena značila, i slično.
- (iii) Koartikulacioni efekat nije jednodimenzionalan problem i ne može se tretirati samo u vremenskom domenu; on se manifestuje i u frekvencijskom i intenzitetskom domenu i nelinearnog je karaktera, tako da se u trodimenzionalnom prostoru vreme-frekvencija-intenzitet trajektorija maksimalne interakcije susednih glasova veoma teško može odrediti pogotovo u spontanom govoru (o ovom problemu biće više reči na kraju rada).

Varijabilnosti u kontinualnom govornom izrazu mogu značajno uticati na akustičku strukturu pojedinačnih glasova (Jovičić, 1999). Kao što je u uvodu naznačeno varijabilnost može biti uzrokovana različitim faktorima: lingvističkim sadržajima, psihomotivnim stanjem govornika, patološkim efektima u artikulaciji glasova, ambijentom u kome se konverzacija odvija i slično. Sa druge strane, artikulacija glasova je jedan fizičko-fiziološki proces nestacionarne prirode koji ima svoje zakonitosti u prelasku iz jednog stanja (generisanja jednog glasa) u drugo stanje (generisanje narednog glasa). U zavisnosti od različitosti ovih stanja, tj. njihovih prostornih udaljenosti u artikulacionom smislu, prelazak iz jednog u drugo stanje može trajati, zbog inercije artikulatora, duže ili kraće uz direktne reperkusije na amplitudsko-spektralne promene u akustičkom govornom signalu. Ovakvo sagledavanje procesa sukcesivnog generisanja glasova dovodi do moguće formulacije *artikulaciono-akustičke distance*. Ovaj pojam je ilustrativno prikazan na slici 2.



Slika 2. Definicija *artikulaciono-akustičke distance* u trodimenzionalnom prostoru sa dimenzijama plozivnost-frikativnost-zvučnost.

U trodimenzionalnom prostoru sa dimenzijama plozivnost-frikativnost-zvučnost ilustrativno su prikazane lokacije pojedinih parova fonetskih podgrupa. Artikulaciono-akustička distanca podrazumeva neposredno rastojanje pojedinih parova tačaka, tj. fonetskih podgrupa, u ovom prostoru. Ako je distanca veća manja je perceptivna konfuzija i lakša je segmentacija. Tako na primer, najteža je segmentacija između vokala i sonanata, a najlakša između vokala i bezvučnih frikativa, afrikata i ploziva, itd. (slika 1).

## OPŠTA NAČELA SEGMENTACIJE

Ova načela su opšteg karaktera i zasnovana su na saznanjima iz oblasti fiziologije artikuladora, fonetske deskripcije glasova i akustičkog modelovanju procesa generisanja glasova.

*Tip segmentacije* - Segmentacija reči se posmatra i tumači sa ekspertskog aspekta, što znači da ekspert koristi prvenstveno svoje znanje za donošenje odluke o granicama segmenata, pri čemu može koristiti razne softverske alate koji mu mogu biti od koristi kao i dostupnu literaturu.

*Segmentacija u vremenu* – Segmentacija reči se vrši u vremenskom domenu, na vremenskoj osi, postavljanjem tačke na govornom signalu na mestu procenjenom za granicu dodira dva fonetska segmenta. Osnova analize u segmentaciji jesu talasni oblik signala, spektrogram i auditivna percepcija.

*Postavljanje tačke segmentacije* – Tačka segmentacije postavlja se na pozitivnoj usponskoj ivici govornog signala na mestu gde govorni signal prolazi kroz liniju nultog intenziteta. Time se obezbeđuje da govorni signal u tački segmentacije polazi iz nultog intenziteta, a ne iz bilo koje tačke signala gde bi se unosio efekat impulsne tranzicije.

*Tačnost segmentacije* – Postavljanje tačke segmentacije podložno je greškama posebno u kontaktu glasova sa jakom koartikulacijom. Određena tolerancija se može dozvoliti na nivou  $\pm 1$  periode osnovne frekvencije glasa (*pitch*-perioda).

*Šta je koartikulacioni segment?* – Koartikulacioni segmenta može se definisati kao *govorni segment u kome se položaj artikulacionih organa menja iz pozicije generisanja prethodnog glasa u poziciju generisanja narednog glasa*. Sa aspekta fonetske segmentacije, kako se tretira u ovom radu, upotreba koartikulacionih segmenata mora biti pažljiva. Zašto? Zato što se koartikulacija kao fiziološka pojava i akustička manifestacija može formulisati kao *asimetrična* i (*uslovno*) *simetrična*.

*Asimetrična koartikulacija* se pojavljuje u kontaktu eksplozivnih i šumnih glasova (plozivi, afrikati i frikativi) sa glasovima *tonske strukture* (vokalima i sonantima). Deo koartikulacije koji pripada eksplozivnom ili šumnom glasu je po prirodi fiziologije artikulacionih organa znatno kraći od dela koartikulacije koji pripada glasovima *tonske strukture*.

*(Uslovno)simetrična koartikulacija* se pojavljuje u kontaktu zvučnih glasova (veoma retko u kontaktu šumnih glasova, na primer /jonosfera/) sa širim tranzicionim zonama kod oba glasa.

*Priroda koartikulacije* – Koartikulacija se može manifestovati na tri načina:

- (i) tranzicijom akustičkih obeležja jednoga glasa u ista akustička obeležja narednog glasa (na primer, tranzicije formantata), tzv. *dinamička tranzicija*,

- (ii) preklapanjem različitih akustičkih obeležja susednih glasova (na primer, nestajanje zvučnosti i nastajanje frikcije), tzv. *mešovita tranzicija*, i
- (iii) trenutna tranzicija akustičkih obeležja jednoga glasa u akustička obeležja narednog glasa (na primer, tranzicije u kontaktu ploziva i vokala), tzv. *eksplozivna tranzicija*.

*Pripadnost koartikulacije* – Tačka segmentacije se nalazi u koartikulacionom segmentu i deli ga na dva dela koja pripadaju glasovima u kontaktu. *Pripadanje* dela koartikulacionog segmenta datom glasu podrazumeva da dati koartikulacioni deo sadrži u većoj meri osobine glasa kome se pripisuje.

*Princip maksimalne brzine tranzicije* – Koartikulacija kod zvučnih glasova tonske strukture, na primer kod vokala i sonanata (primer na slici 1), je veoma izražena, spora i zahvata širu vremensko-spektralnu zonu u spektrogramu. To ukazuje na „duboko“ prožimanje glasova. Tranzicija položaja artikulacionih organa ima svoju trajektoriju koja se manifestuje u spektralnom domenu, na primer tranzicijom formanata. Tranzicija se odvija različitom brzinom i najveća je u središnjem delu tranzicionog segmenta. Na mestu najveće brzine tranzicije odvija se najsnažnije prožimanje glasova i može se prihvatiti kao mesto segmentacije.

*Granice koartikulacionog segmenta* – Nastanak ili prestanak koartikulacije je teško odrediti pošto artikulacioni organi u svom delovanju čine kontinualan sistem. Pošto su i sami glasovi u koartikulacionom kontaktu u nestabilnom režimu generisanja, za nastanak/prestanak koartikulacije može se prihvatiti kriterijum odstupanja trenda promene akustičkih obeležja u očekivanoj zoni koartikulacije. Na primer, kod dvoglasa VC (vokal-konsonant) drugi formant može imati u zoni vokala blagi opadajući nagib, ali kada ovaj trend počne naglo da se menja sa približavanjem konsonantu, da opada ili raste, u tom trenutku treba postaviti granicu koartikulacionog segmenta. Greška u postavljanju ove granice jeste individualne prirode, ali ako je analitičar konzistentan u svom pristupu sistematska greška se može uračunati u dalje postupke analize.

*Postavljanje granice segmentacije* – Kod asimetrične koartikulacije granica segmentacije se postavlja na mestu jasnog amplitudskog ili spektralnog diskontinuiteta, dok se kod (uslovno) simetrične koartikulacije granica segmentacije postavlja, u slučaju *dinamičke koartikulacije*, na mestu maksimalne brzine tranzicije koartikulacije ili, u slučaju *mešovite koartikulacije*, na sredini koartikulacionog segmenta.

*Artikulaciono-akustička distanca glasova* – Definisana je slikom 2 i pripadajućim tekstom.

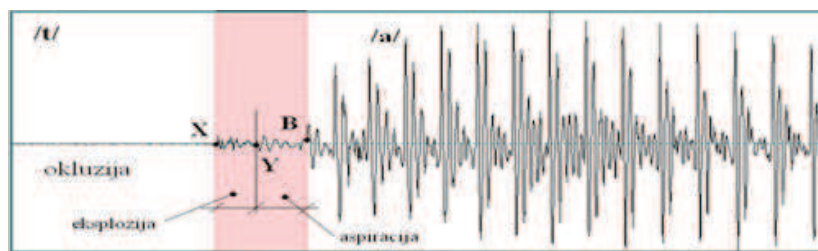
## PRIMENA OPŠTIH NAČELA SEGMENTACIJE

U narednim razmatranjima posmatraće se položaj glasa u okviru reči, pa će segmentacijom biti obuhvaćeno i određivanje inicijalne i finalne granice reči.

### Određivanje inicijalne granice reči

*Slučaj bezvučnih ploziva i afrikata* – Bezvučni plozivi /p, t, k/ i afrikati /c, ć, č/ u inicijalnoj poziciji u reči imaju okluziju kod koje ne postoji inicijalna granica. Ukoliko

je potrebno definisati inicijalnu granicu okluzije, tada se inicijalna granica može postaviti ispred pojave eksplozije ili frikcije na rastojanju koje odgovara srednjem trajanju okluzije odgovarajućeg glasa. Na slici 3 prikazan je primer dvoglasa /ta/ u reči “tata” sa subsegmentima ploziva /t/, eksplozija i aspiracija koji su zajedno u literaturi poznati pod nazivom VOT (*Voice-Onset-Time*), i tačkom B nultog intenziteta na usponskoj ivici prvog impulsa koji pripada vokalu /a/. U ovom slučaju tačka B nedvosmisleno predstavlja granicu između eksplozivne strukture glasa /t/ i kvaziperiodične strukture glasa /a/.



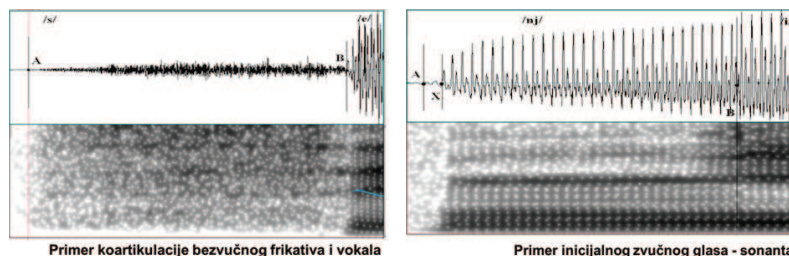
**Primer inicijalnog bezvučnog ploziva u spoju sa vokalom**

Slika 3. Subsegmenti ploziva /t/ i tačka segmentacije B između glasova /t/ i /a/.

*Slučaj bezvučnih frikativa* – Bezvučni frikativi /f, h, s, š/ su šumnog karaktera i nastaju sporijim porastom intenziteta šumnosti. Granica nastanka šumnosti se jasno uočava i u intenzitetskom i u spektralnom domenu, na talasnom obliku signala i spektrogramu, respektivno (primer frikativa /s/ u reči “seka” na slici 4). Postupak postavljanja inicijalne granice zavisi od kvaliteta snimka, odnosno od prisutnih smetnji u snimku.

Ako nema smetnji ili su jako malog intenziteta, tada se preciznija granica može postaviti u intenzitetskom domenu, na talasnom obliku signala, na mestu početka uvećanja intenziteta signala. Ova granica u spektralnom domenu ne mora biti jasno izražena.

Ako su smetnje intenzivnije tada dolazi do maskiranja govornog signala i inicijalna granica u vremenskom domenu se teško može odrediti. Pouzdanija procena inicijalne granice se može napraviti u spektralnom domenu zbog realne pretpostavke da spektralni sastavi smetnje i frikcije navedenih frikativa nisu isti. U tom slučaju inicijalnu granicu treba postaviti na mestu pojave spektralnog sadržaja frikcije na spektrogramu.



Slika 4. Postavljanje inicijalne granice bezvučnog frikativa /s/, tačka A, i nazala /nj/, tačka X.

*Slučaj zvučnih konsonanata* – Svi zvučni glasovi: plozivi /b, d, g/, frikativi /z, ž/, afrikati /đ, dž/ i sonanti /m, n, nj, v, r, l, lj, j/, imaju jasno izraženu zvučnost. Nastajanje zvučnosti počinje inicijalnim segmentom najčešće u trajanju jedne do dve periode osnovne frekvencije glasa, *pitch*-periode (segment A-X u primeru nazala /nj/ u reči “njiva” na slici 4). U ovom segmentu artikulacioni organi, pre svega glasnice, iz stanja mirovanja uvode se u režim generisanja zvučnosti, odnosno režim fonacije. Zbog toga ovaj segment nema fonemsku funkciju te se inicijalna granica postavlja na mestu pojave prvog glotalnog impulsa (tačka X na slici 4) kada se ostvari normalan režim generisanja zvučnosti. „Normalan režim“ se jasno uočava po talasnoj formi prve kvalitetno generisane *pitch*-periode, kao i u spektralnom domenu po kompletiranom spektralnom sadržaju zvučnosti datog glasa.

#### Određivanje finalne granice reči

*Slučaj finalnih vokala* – S obzirom da finalni vokali mogu biti u značajnoj meri obezvučeni mogu se posmatrati dva slučaja: zvučna varijanta finalnog vokala (sa *čvrstom artikulacijom*) i obezvučena varijanta finalnog vokala (sa *nepreciznom artikulacijom*).

*Zvučna varijanta finalnog vokala* – U ovom slučaju artikulacija vokala je čvrsta i kao takva održava se do kraja njegove produkcije. Dva osnovna kriterijuma za procenu finalne granice reči, tj. kraja artikulacije vokala, su:

- asimptotski pad anvelope finalnog vokala, i
- nestajanje formanta u spektrogramu; s obzirom na spektralnu karakteristiku govora prvo nestaju najviši formanti, a najduže se, po pravilu, zadržava prvi formant (treba imati u vidu da formanti od prvog ka višim formantima ne opadaju linearno po energiji, što bi se očekivalo imajući u vidu dugovremeni spektar govornog signala, već da njihov energetska odnos zavisi od trenutne produkcije datog vokala).

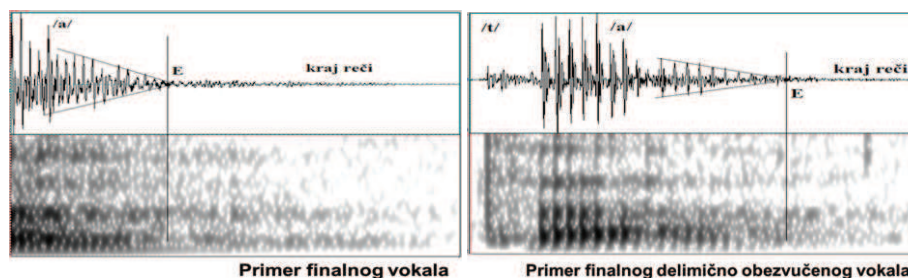
Prvi kriterijum ukazuje na opadanje intenziteta signala vokala ka nultoj vrednosti, kada prestaje vazдушna struja kroz glotis i kada prestaju vibracije glasnice. Ovaj fiziološki proces je veoma kompleksan. Naime, glasnice kao biološko-mehanički element imaju svoju masu i inerciju i na kraju fonacije prestaju sa vibracijama pri dovoljno malom protoku vazduha kroz glotis, odnosno malom akustičkom pritisku. Ovaj period artikulacije se naziva labava artikulacija. Posle tog trenutka protok vazduha stvara šumnu strukturu koja nestaje tek sa potpunim prestankom izdisaja. Šumna struja prolaskom kroz vokalni trakt generiše šaptavu verziju vokala, ukoliko je struktura vokalnih šupljina ostala nepromenjena, ili formu glasa bez fonetskog značaja, ukoliko su artikulacioni organi poprimili nedefinisane formu.

Drugi kriterijum proizilazi iz činjenice da trougaoni glotalni impulsi, bogati harmonicima koji formiraju formante, sa tranzicijom ka prestanku vibriranja glasnice prelaze u kvazi-sinusoidalnu formu sa malim sadržajem harmonika. Posledica ove promene je nestajanje zvučne strukture formanta, ostaje šumni sadržaj formanta i rezidual laringalne zvučnosti. Ovakav sadržaj generisanog govornog signala nema više fonetski značaj, ali označava kraj vokalne aktivnosti u fazi ekspirijuma.

Finalnu granicu treba postaviti na mestu gde se susreću nestanak poslednjeg formanta, najčešće prvog formanta, i asimptotski pad anvelope govornog signala (primer fi-



nalnog /a/ u reči “šuma” na slici 5). Nakon ove granice ostaje smirujući signal inercijalne manifestacije celokupnog vokalnog mehanizma, bez fonetskog značaja. Bitna činjenica je da se njegovim odstranjivanjem iz signala cele reči nimalo ne narušava percepcija ni finalnog vokala ni cele reči.



Slika 5. Finalna granica vokala /a/ i delimično obezvučenog vokala /a/.

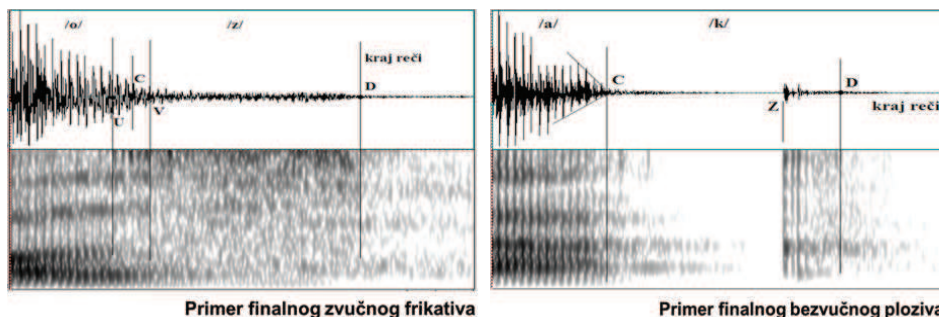
*Obezvučena varijanta finalnog vokala* – Ukoliko artikulacija govornog izraza nije čvrsto strukturirana, imamo labavu artikulaciju, a njene manifestacije u govornom signalu su:

- značajno smanjenje intenziteta govornog signala, posebno na kraju govornog izraza u zoni finalnog glasa,
- redukcija glotalne aktivnosti do potpunog prekida i zamena zvučnih struktura bezvučnim,
- neregularnosti (nestabilnosti i nesigurnosti) u artikulacionoj aktivnosti koje su jasno uočljive u finalnom delu spektrograma.

Opšta slika obezvučenog vokala jeste izrazito prisustvo šumnih struktura, a manje prisustvo zvučnih struktura, i opšta nestabilnost svih struktura. To znači da se percepcija obezvučenog vokala doživljava kao percepcija vokala u šapatu. Formantna struktura spektra vokala i dalje postoji ali su, pogotovo viši formanti, pretežno šumnog karaktera. Odnos tonske i šumne strukture zavisi od stepena obezvučenosti vokala.

Finalna granica obezvučenih vokala se postavlja na istim principima kao i kod zvučnih vokala. Kod određivanja ove granice treba biti obazriv s obzirom da proces nestajanja vokala zbog labave artikulacije najčešće nije kontinualne prirode, tako da se u artikulaciji mogu pojaviti, zbog slabe kontrole artikulatora, prekidi u spektralnim strukturama formanta (što se uočava na slici 5 u finalnom vokalu /a/ u reči “tata”). Zbog toga treba pratiti opšti trend opadanja amplitudsko-spektralnih struktura do potpunog nestanka prvog formanta.

*Slučaj finalnih zvučnih konsonanata* - Kod zvučnih konsonanata u finalnoj poziciji veoma je prisutan efekat obezvučavanja. Ovaj efekat je posebno izražen kod zvučnih frikativa i afrikata gde, zbog obezvučavanja, zvučnom glasu preostaju primarne karakteristike njegovog bezvučnog parnjaka. Na slici 6 prikazan je primer finalnog zvučnog frikativa /z/ u reči “voz”. U spektrogramu se jasno uočava nestanak zvučnosti već na samom početku generisanja glasa /z/, nakon tačke V, dok preostala šumna struktura u potpunosti odgovara glasu /s/. Finalna granica, tačka D, se teško može definisati na talasnom obliku govornog signala ali je veoma jasno izražena u spektrogramu.

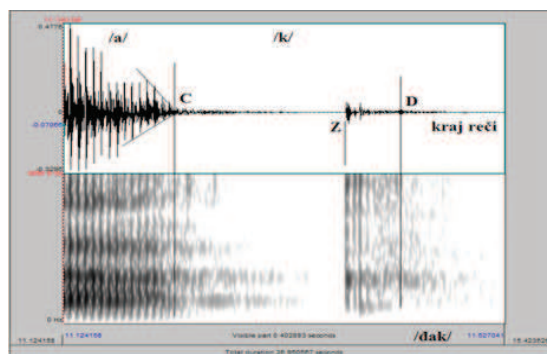


Slika 6. Finalna granica obezvučenog frikativa /z/ i bezvučnog ploziva /k/.

U slučaju preostalih zvučnih konsonanata, nazala i aproksimanata, finalna granica se postavlja po istim kriterijumima kao za finalne vokale.

*Slučaj finalnih bezvučnih frikativa i afrikata* – Određivanje finalne granice ovih glasova se zasniva na sličnim kriterijumima kao i određivanje inicijalne granice. Finalna granica se postavlja na mestu određenom sa dva kriterijuma: (1) asimptotskim opadanjem amplitude signala frikativa odnosno afrikata, što je labaviji kriterijum s obzirom na moguće prisustvo šuma, i (2) naglim prestankom frikcije vidljive u spektrogramu, što je pouzdaniji kriterijum. U spektrogramu treba uočiti razliku između kraja frikcije i šuma ekspirijuma posle finalne granice, kao posledice potpune relaksacije artikulacionih organa. Ovi detalji su uočljivi i na slici 6 kod obezvučenog glasa /z/ za koga je konstatovana potpuna transformacija u frikativ /s/.

*Slučaj finalnih ploziva* - Bez obzira da li se radi o zvučnom ili bezvučnom plozivu, u finalnoj poziciji u reči svi ovi glasovi završavaju se eksplozivnim segmentom. Finalna granica se postavlja na kraju aspirovanog segmenta posle koga sledi šum ekspirijuma. Diskriminaciju između aspirovanog dela eksplozije ploziva i ekspirijuma je teško ili čak nemoguće učiniti bez eksplicitnog indikatora u smislu spektralne ili amplitudske neregularnosti koje se pripisuju ekspirijumu. U primeru na slici 7 tačka D predstavlja finalnu granicu bezvučnog ploziva /k/ u reči /đak/.



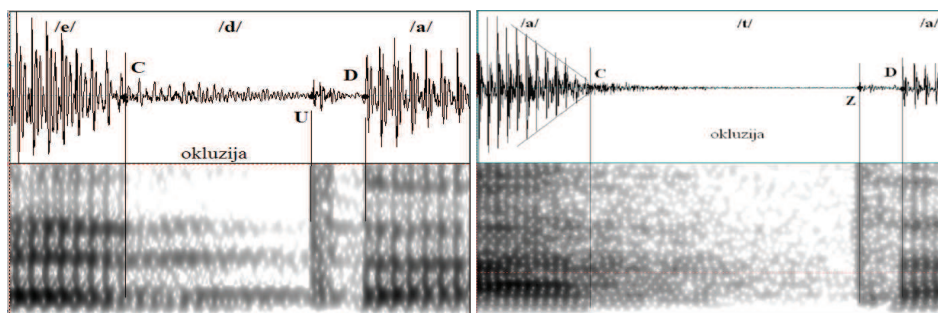
### Određivanje granice ploziva i vokala

*Spoj ploziv-vokal* – Kontakt ove dve fonetske grupe pripada veoma čvrstoj, tj. preciznoj artikulaciji. Samo stvaranje pregrade u vokalnom traktu, da bi se uopšte produkovao ploziv, najavljuje eksplozivnu prirodu nastajanja vokala koji sledi. U oba slučaja zvučnog i bezzvučnog ploziva njihov kontakt sa vokalom se nalazi u tački dodira VOT segmenta i prvog glotalnog impulsa u vokalu. Na slici 8 u oba primera, kod zvučnog ploziva /d/ u reči “deda” i kod bezzvučnog ploziva /t/ u reči “tata”, tačka segmentacije sa vokalom /a/ je tačka D.

*Spoj vokal-ploziv* – Ovaj spoj ne mora biti jasno profilisan iz tri razloga: (i) vokal se ne može plozivno završiti, (ii) postoji inercija artikulacionih organa u prelazu sa zvučnosti na okluziju, i (iii) okluzija često nije potpuna (pitanje kvaliteta okluzije).

Ako je kod bezzvučnih ploziva okluzija kvalitetna, tada se tačka segmentacije postavlja iza poslednjeg generisanog glotalnog impulsa kojim se završava generisanje vokala (primer dvoglasa /at/ u reči “tata” i tačke C na slici 8). Ako okluzija nije kvalitetna, tada se laringalne vibracije mogu produžiti u okluziju i sa opadanjem intenziteta nastaviti skoro do VOT segmenta (slično manifestaciji zvučnosti kod zvučnih ploziva). U amplitudskom domenu prelazak sa vokala na laringalne vibracije u okluziji ne mora biti izražen, ali u spektralnom domenu ova razlika je jasno naznačena završetkom formanta vokala. Tačka segmentacije se postavlja na mestu jasno izražene amplitudske razlike između vokala i okluzije ili na mestu završetka poslednjeg formanta, najčešće prvog formanta (napomena: formanti se najčešće ne završavaju sinhrono, najpre u spektrogramu nestaju viši formanti kao slabiji po intenzitetu, a na kraju drugi ili/i prvi formant kao najintenzivniji formanti – ova pojava je posledica oblika spektra govornog signala).

Kod zvučnih ploziva laringalne vibracije postoje kao inherentni deo ploziva i situacija je ista kao kod nekvalitetne okluzije bezzvučnih ploziva (prethodno opisana). U ovom slučaju intenzitetska razlika između vokala i zvučnosti okluzije je manja pa se tačka segmentacije pouzdanije može odrediti u spektralnom domenu (primer dvoglasa /ed/ u reči “deda” na slici 8, i tačka segmentacije C).



Slika 8. Primeri segmentacije ploziva u intervokalskom okruženju.

### Određivanje granice frikativa i vokala

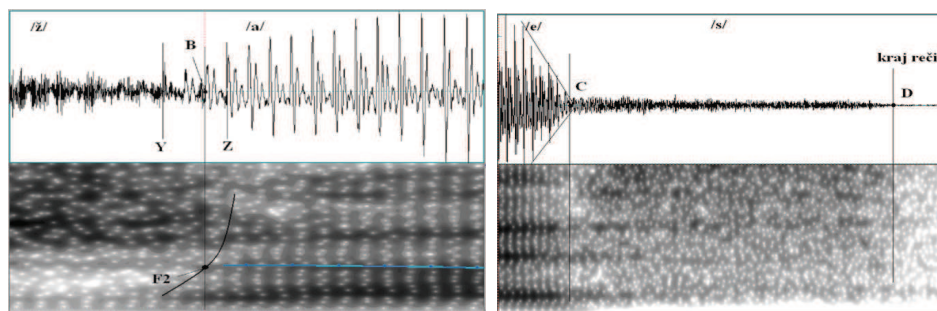
*Spoj frikativ-vokal* – Kontakt ove dve fonetske grupe pripada čvrstoj, tj. preciznoj artikulaciji. Artikulacija je potpuno različite prirode, posebno kada su u pitanju bezzvučni

frikativi, tako da se artikulacija glasa iz jedne fonetske grupe mora u potpunosti završiti da bi se započela artikulacija glasa iz druge fonetske grupe. U takvim okolnostima koartikulacija je izuzetno slaba ili je uopšte nema.

U slučaju bezvučnog frikativa kontakt frikcije i zvučnosti je veoma precizan i vidljiv i u talasnom obliku signala i u spektrogramu. Koartikulacija je minimalna, skoro da je nema, a tačka segmentacije se postavlja na usponskoj ivici prvog glotalnog impulsa vokala. Primer ovakvog slučaja je prikazan na slici 4 za frikativ /s/ u reči “seka” gde je tačka segmentacije označena sa B.

Segmentacija zvučnog frikativa je nešto složenija jer postoji kontinuitet zvučnosti bez obzira što je zvučnost u frikativu znatno intenzitetski slabija od zvučnosti vokala. Zbog toga postoji koartikulacioni segment (primer na slici 9, spoj frikativa /ž/ i vokala /a/ u reči “žaba” i koartikulacioni segment Y-Z) u kome zvučnost frikativa prerasta u zvučnost vokala i u kome frikcija slabi i nestaje, odnosno u kome se preklapaju pre svega prvi formant vokala i završna šumnost frikativa. Tačka segmentacije se postavlja u sredini koartikulacionog segmenta na usponskoj ivici najbližeg glotalnog impulsa.

*Treba uočiti da separacija frikativa i vokala nije linearna funkcija u trodimenzionalnom prostoru (intenzitet-vreme-frekvencija), zbog čega nije jednostavno postaviti tačku segmentacije samo na vremenskoj osi. Ovaj stav važi za bilo koji koartikulacioni segment i o njemu će biti više reči u sledećem segmentu ovog rada.*



Slika 9. Primeri segmentacije zvučnog frikativa /ž/ i vokala /a/ u reči “žaba” (tačka B) i vokala /e/ i bezvučnog frikativa /s/ u reči “fes” (tačka C).

*Spoj vokal-frikativ* – U slučaju bezvučnog frikativa spoj vokala i frikcije je jasno izražen prestankom svih formantnih struktura vokala i nastankom intenzivne frikcije. Tačka segmentacije se postavlja iza poslednjeg uočljivog laringalnog impulsa (primer vokala /e/ i bezvučnog frikativa /s/ u reči “fes” gde je tačka segmentacije označena sa C, slika 9). U segmentu frikativa moguća je pojava šumnih spektralnih koncentrata na nivou formantata prethodnog vokala što je posledica inercije artikulacionih organa, što nije od primarnog značaja za percepciju frikcije.

Kod zvučnih frikativa, na primer u spoju vokala /o/ i zvučnog frikativa /z/ na slici 6, situacija je kompleksnija iz osnovnog razloga što se radi o zvučnim glasovima. U ovom slučaju pojavljuje se koartikulacioni segment U-V u kome se preklapaju završeci najnižih formantata vokala i nastanak frikcije u najvišem delu spektra (mešovita tranzicija). Ove karakteristike se jasno uočavaju u spektrogramu gde se mogu definisati granice koarti-

kulacionog segmenta. Tačka segmentacije C se određuje na bazi usvojenog principa na sredini koartikulacionog segmenta.

NAPOMENA: Generalno, u zavisnosti od individualne produkcije ovih glasovnih spojeva, posebno u kontinuiranom govoru, princip sredine koartikulacionog segmenta može se uvek primeniti.

#### Određivanje granice afrikata i vokala

Činjenica da je artikulacija afrikata kombinacija plozivnosti i frikcije u značajnoj meri pojednostavljuje postupak segmentacije afrikata i vokala. To znači da je kontakt vokal-afrikat isti kao kontakt vokal-ploziv, a da je kontakt afrikat-vokal isti kao kontakt frikativ- vokal, pa se i segmentacija afrikata i vokala vrši prema prethodno opisanim kriterijumima za plozive i frikative.

#### Određivanje granice nazala i vokala

Iako su nazali konsonanti (sonanti) u kojima postoji kombinacija tonske i šumne strukture, njihova segmentacija u spoju sa vokalima, pa i drugim konsonantima, je veoma olakšana zahvaljujući njihovoj plozivnoj prirodi generisanja. Zatvaranje usne šupljine i otvaranje nosne šupljine, kao i obrnuto, je nagli proces koji se odvija na nivou jednog laringalnog impulsa, tako da se koartikulacioni proces odvija veoma brzo u okviru jedne *pitch*-periode. Tačka segmentacije se postavlja na usponskoj ivici prvog glotalnog impulsa koji pripada vokalu odnosno nazalu.

NAPOMENA: Plozivnost nazala u kontinualnom govoru, brzom govoru ili govoru bez precizne artikulacije može biti veoma slabo izražena ili čak može izostati. U tom slučaju se pojavljuje koartikulacioni segment u oba spoja sa vokalom i treba primeniti kriterijume segmentacije kao za ostale sonante.

#### Određivanje granice sonanata i vokala

Sonanti su po preovlađujućoj tonskoj strukturu najbliži vokalima i, prema ranijem šematskom prikazu na slici 2, imaju najmanju artikulaciono-akustičku distancu u odnosu na vokale. To znači da je njihovo prožimanje sa vokalima najveće, a samim tim i segment koartikulacije je najveći (videti spektrogram reči “jaje” na slici 1).

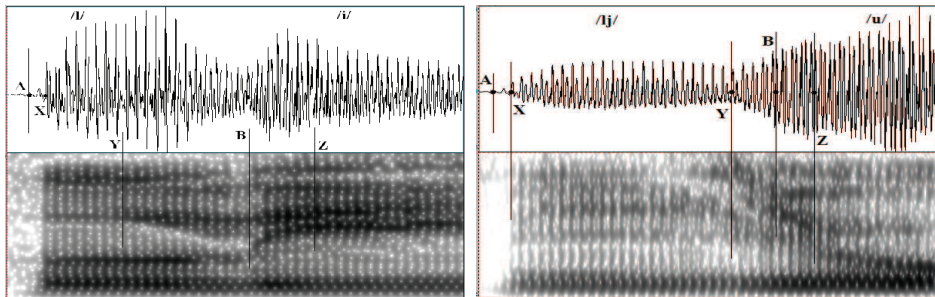
U vezi sa koartikulacijom kod sonanata treba naglasiti i sledeće karakteristike:

- trajanje segmenta koartikulacije je duže kada je vokal ispred sonanta (ova osobina je izražena i u spojevima kod ostalih zvučnih konsonanata),
- trajanje segmenta koartikulacije zavisi od pozicije formantnih struktura i sonanata i vokala u spoju; što su pozicije udaljenije, trajanje koartikulacije je duže.

U ovoj analizi treba uzeti u obzir mogućnost:

- da koartikulacija može biti asimetrična kao i uslovno simetrična, primeri na slici 10 za laterale /l/ u reči “lice” i /lj/ u reči “ljudi” gde su koartikulacioni segmenti označeni sa Y-Z a tačka segmentacije sa B,
- da se u koartikulacionom segmentu laterala može pojaviti jasan diskontinuitet kao posledica odvajanja jezika od nepca (kao u izgovoru laterala /l/) koji neposredno ukazuje na tačku segmentacije, i
- da realizacija vibranta /r/ može biti veoma različita od slučaja jasne forme od nekoliko akustičkih impulsa do slučaja bez akustičkog impulsa, odnosno akustičkog impulsa integrisanog sa glasom u kontaktu.

Bez obzira koji je spoj u pitanju vokal-sonant ili sonant-vokal tačka segmentacije se postavlja na mestu najbrže tranzicije spektralnih struktura.



Slika 10. Primeri segmentacije laterala i vokala.

## IZAZOVI ZA BUDUĆA ISTRAŽIVANJA

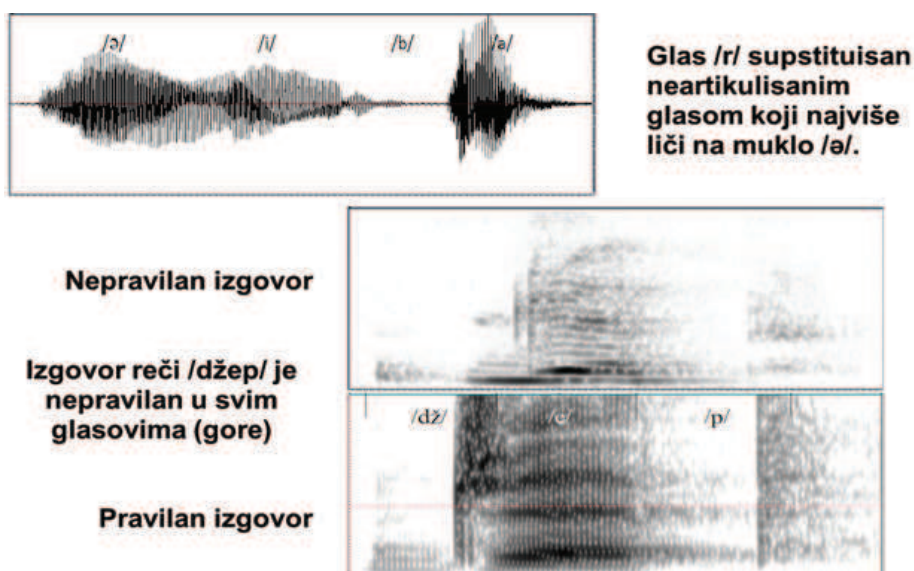
Prethodnim primerima demonstrirana je primena usvojenih principa segmentacije glasova imajući u vidu sekvencijalnu prirodu produkcije glasova u kontinualnom govoru. Takvo gledište je nametnulo razmatranje segmentacije glasova u vremenskom domenu. Detaljan uvid u artikulaciono-akustičku prirodu glasovnih spojeva izbacio je u prvi plan koartikulacioni segment, kao jedan od ključnih prozodijskih elemenata koji obezbeđuje prirodnost verbalne komunikacije, u kome je segmentacija glasova locirana. Kompleksnosti ovom problemu doprinose i sledeća pitanja.

*I. Sva dosadašnja razmatranja su se odnosila na standardan (tipičan) govor ali se postavlja pitanje da li je moguće i kako segmentirati nestandardan govor kao što su emotivni govor; govor sa odstupanjima od tipičnog govora (patološki govor) ili govor šapatom?*

U odgovoru na ovo pitanje možda je najteži problem - patološki govor. Na slici 11 prikazana su dva primera: prvi se odnosi na supstituciju vibranta /r/ u reči “riba” redukovanim glasom /ə/, a drugi na potpuno nepravilan izgovor svih glasova u reči “džep”.

Najveći problem kada je u pitanju govor sa odstupanjima od tipičnog govora jeste nepredvidivost artikulacije, tj. artikulacije bez fonetsko-fonološke osnove koja se može manifestovati u bezbroj varijanti. Ako je u pitanju kontakt glasa sa artikulacionim odstupanjem i normalnog glasa, kao što je prvi slučaj u reči “aiba”, tada je poznavajući karakteristike normalnog glasa, u ovom slučaju vokala /i/, segmentacija moguća. Ali ako je u pitanju nepravilna artikulacija glasova u klasteru, kao u primeru “džep”, tada se veoma teško može izvršiti pouzdana segmentacija, ako je uopšte i moguća.

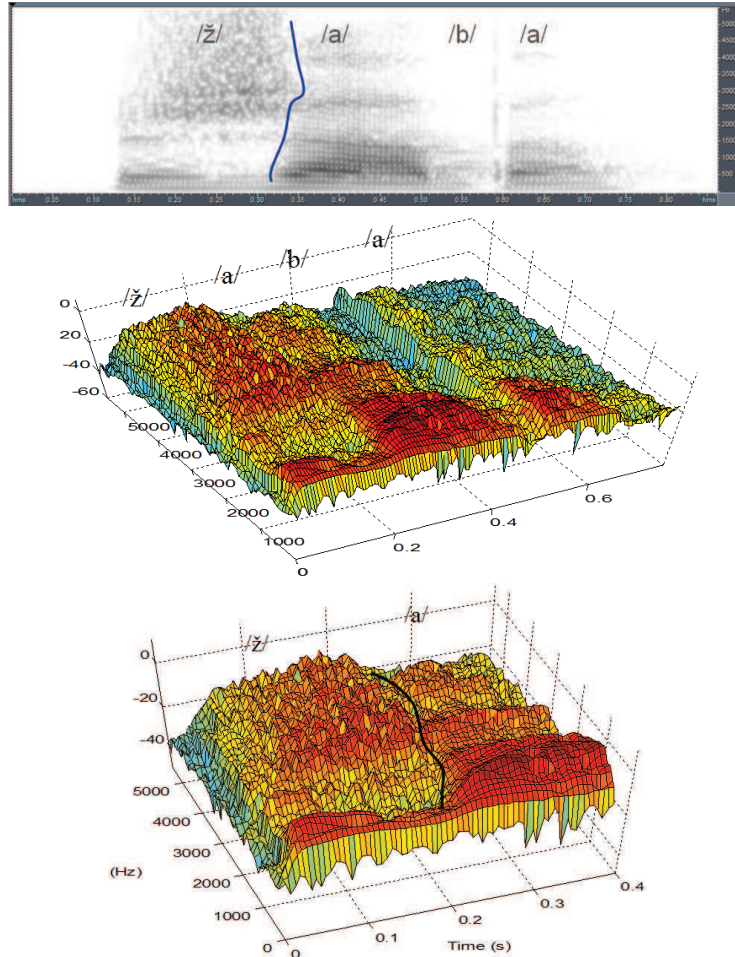
U ovakvim slučajevima neophodno je interdisciplinarno sagledavanje problema segmentacije, odnosno u proceni segmentacije moraju biti upotrebljena znanja iz oblasti logopedije, fonetike, fiziologije i akustike (Šarić et al., 2007; Jovičić et al., 2010; Bilibajkić et al., 2016).



Slika 10. Primeri nepravilnih izgovora pojedinih glasova.

II. U više prethodno prikazanih primera segmentacije susednih glasova ukazano je na postojanje tzv. *mešovite tranzicije* tokom koje dolazi do mešanja ili preklapanja u vremenskom domenu akustičkih obeležja susednih glasova (primer na slici 9 kod spoja glasova /ž/ i /a/ u reči “žaba”). Postavlja se pitanje da li je opravdano tretirati segmentaciju samo u vremenskom domenu?

Tranzicija glasa /ž/ u glas /a/, u pomenutoj reči “žaba”, posebno je markirana u spektrogramu na slici 11. Linijom nepravilnog oblika napravljeno je razgraničenje između frikativne strukture frikativa /ž/ i kvaziperiodične formantne strukture vokala /a/. U donjem redu slike 11 levo prikazan je spektrogram u trodimenzionalnoj (3D) formi za celu reč “žaba” i u desnom delu 3D spektrogram izdvojenog dvoglasa /ža/.



Slika 11. Primer u prilog 3D segmentacije.

U ovoj reljefnoj strukturi jasno se uočavaju intenzitetsko-spektralno-vremenski detalji oba glasa, a linijom je naznačeno njihovo strukturalno razgraničenje, tj. segmentacija. Očigledno je da se pitanje segmentacije ne može rešiti samo u jednom domenu, vremenskom ili intenzitetskom ili spektralnom, već da se mora primeniti integralni pristup u formulaciji nekog algoritma za segmentaciju. Prvi korak u realizaciji algoritma je svakako definisanje kriterijuma za segmentaciju. U prvom viđenju, imajući u vidu 3D spektrogram na slici 11, dva kriterijuma mogu biti polazna osnova u razvoju algoritma:

- (i) maksimum entropije (minimum energije) u zoni segmentacije, i
- (ii) maksimum diferencijacije susednih regiona glasova u kontaktu.

Dalji postupci u realizaciji jednog ovakvog algoritma zahtevaju upotrebu matematičko-softverskih znanja (Huang et al., 2001), što izlazi van okvira ovog rada.



### III. Prethodna diskusija je upravo nametnula pitanje da li se segmentacija može vršiti automatski - računarski?

Svakako da je moguća i postoje određena rešenja koja su razvijena u različitim aplikacijama u okviru tzv. govornih tehnologija (Sarma and Sarma, 2014). U ovom cilju je neophodno poznavati veoma dobro matematičke algoritme u obradi govornog signala i, naravno, njihovu softversku realizaciju. Ako je već tako, onda je viđenje 3D segmentacije moguće preformulisati u *multidimenzionalnu segmentaciju*, pošto se u obradi govornog signala veoma mnogo koriste parametarska obeležja, obeležja matematičke prirode (Benesty et al., 2008). Naravno, ovo je već stvar budućih istraživanja.

### Literatura

- Benesty, J., Sondhi, M.M. and Huang, Y. (2008). *Handbook of speech processing*. Springer-Verlag, Berlin.
- Bilibajkić, R., Šarić, Z. and Jovičić, S.T. (2011). Auditory-model based speech segmentation on subphonemic segments. In *Proceedings of Forum Acusticum 2011*, Aalborg, Denmark, 55-60.
- Bilibajkić, R., Šarić, Z., Jovičić, S.T., Punišić, S. and Subotić, M. (2016). Automatic detection of stridence in speech using the auditory model. *Computer Speech and Language*, 36, 122-135.
- Glass, R.J. and Zue, V.W. (1988). Multi-level acoustic segmentation of continuous speech. In *Proceedings of ICASSP*, 429-432.
- Hatazaki, K., Komori, Y., Kawabata, T. and Shikano, K. (1989). Phoneme segmentation using spectrogram reading knowledge. In *Proceedings of CASSP*, 393-396.
- Huang, X., Acero, A. and Hon, H.W. (2001). *Spoken language processing: A guide to theory, algorithm and system development*. Prentice-Hall, New Jersey.
- Jovičić, S.T., Kašić, Z. and Punišić, S. (2010). Production and perception of distortion in word-initial friction duration. *Journal of Communication Disorders*, 43(5), 335-346.
- Jovičić, S.T. (1999). *Govorna komunikacija: fiziologija, psihoakustika i percepcija*. Nauka, Beograd.
- Kašić, Z. (1980). Glasovne promene u proklizi u srpskohrvatskom jeziku. *Naš jezik*, XXIV/4-5, 217-246.
- Kašić, Z. (1985). Glasovne promene u enklizi. *Naš jezik*, XXVI//4-5, 228-233.
- Kašić, Z. (1990). Sandhi i neutralizacija distinktivnih obeležja. *Književnost i Jezik*, 1, 71-73.
- Kašić, Z. (1997). Promene glasova uslovljene ritamsko-intonacionom organizacijom govora. *Beogradska defektološka škola*, 1, 77-82.
- Kašić, Z. (2000). *Segmentna i suprasegmentna organizovanost govora*. U knjizi: Golubović S., Kašić Z.: Segmentna i suprasegmentna organizovanost govora i poremećaji fluentnosti. Beograd: Društvo defektologa Jugoslavije, 9-62.
- Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. *Journal of Acoustical Society of America*, 58(4): 880-883.
- Sarma, M. and Sarma, K.K. (2014). *Phoneme-based speech segmentation using hybrid soft computing framework*. Springer, New Delhi.
- Shriberg, E., Cuendet, S., Favre, B., Fung, J. and Hakkani-Tur, D. (2008). Prosodic similarities of dialog act boundaries across speaking styles. In *Linguistic Patterns in Spontaneous Speech* (Language and Linguistics Monograph Series), S.-C. Tseng, Ed. Institute of Linguistics, Academia Sinica, Taipei.
- Simeon, R. (1969). *Enciklopedijski rječnik lingvističkih naziva*, I-II. Zagreb: Matica Hrvatska.

Šarić, Z., Jovičić, S.T. and Bilibajkić, R. (2007). Speech segmentation on subphonemic segments aimed to the speech pathology analysis. *19<sup>th</sup> International Congress on Acoustics, ICA 2007*, paper CAS-03-040, Madrid.

**Slobodan Jovičić**  
**Zorka Kašić**

**SOUNDS SEGMENTATION IN SPEECH EXPRESSION:  
PROBLEMS AND PRINCIPLES**

**Summary:** This paper deals with an issue of sounds segmentation in continuous speech in which, between two breathing pauses, all sounds (independently from the fact whether they belong to different words) make contact mutually. The general principles of segmentation are conveyed, based on the knowledge from the field of articulator physiology, phonetic sounds description and acoustic modeling of generating sounds process. It discussed possible formulation of articulation-acoustic distance in three-dimensional space.

According to the given examples, suggestions for segmentation are given, segmentation of certain classification of sounds groups in mutual contact. As the most important problems in the process of segmentation there were: determining the beginning of voiceless plosives in initial position, various types coarticulation effects in vowel contacts with consonants and sonants, as producing sounds voiceless in absolute final position in phonetic phrase, which leads to complexity in determining the end of pronunciation. The suggestions may be useful as bases for further research of the given views, and as an incentive for the development of mathematical algorithms with application of information technologies in speech communication. At the end of this paper the fact that a question of segmentation should be viewed, not only in time dimension but in three-dimensional space time-spectrum-intensity.

**Key words:** sounds segmentation, segmentation principles, coarticulation, three-dimensional segmentation, classification of a group of sounds.