UNIVERZITET U BEOGRADU

ELEKTROTEHNIČKI FAKULTET

Ružica Bilibajkić

PREPOZNAVANJE ARTIKULACIONO-AKUSTIČKIH ODSTUPANJA GLASOVA U PATOLOŠKOM GOVORU

doktorska disertacija

Beograd, 2016

UNIVERSITY OF BELGRADE

SCHOOL OF ELECTRICAL ENGINEERING

Ružica Bilibajkić

DETECTION OF ARTICULATORY-ACOUSTIC DEVIATIONS IN PATHOLOGICAL SPEECH

Doctoral Dissertation

Belgrade, 2016

PODACI O MENTORU I ČLANOVIMA KOMISIJE:

Mentor:

dr Dragana Šumarac Pavlović, vanredni profesor, Univerzitet u Beogradu – Elektrotehnički fakultet

Članovi komisije:

dr Dragana Šumarac Pavlović, vanredni profesor, Univerzitet u Beogradu – Elektrotehnički fakultet

dr Miomir Mijić, redovni profesor, Univerzitet u Beogradu – Elektrotehnički fakultet

dr Zoran Šarić, naučni savetnik, Centar za Unapređenje Životnih Aktivnosti, Beograd

dr Jelena Ćertić, docent, Univerzitet u Beogradu – Elektrotehnički fakultet

datum odbrane: _____

Zahvalnica

Hvala prof. Slobodanu T. Jovičiću, dr. Zoranu Šariću, prof. Dragani Šumarac Pavlović i prof. Miomiru Mijiću na inspiraciji i pruženoj prilici da se bavim naučno-istraživačkim radom.

Hvala kolegama na nesebičnoj pomoći.

Posebnu zahvalnost dugujem svojoj porodici na svoj pruženoj podršci proteklih godina. Njima je ova disertacija i posvećena.

Naslov: Prepoznavanje artikulaciono-akustičkih odstupanja glasova u patološkom govoru

Rezime

U kliničkoj praksi procena artikulacionih poremećaja uglavnom se vrši pomoću subjektivnih mera i metoda, audio vizuelnom procenim od strane obučenih eksperata-logopeda. U tu svrhu postoji određeni broj standardnih testova i procedura na osnovu kojih se daje kvalitativna i kvantitativna ocena patološkog izgovora. Ovakav pristup podložan je uticaju faktora koji potiču od ispitivača i uslova ispitivanja, a koji nisu posledica patološkog izgovora. Postojanje automatskog sistema za procenu patologije u govoru, u prvom redu bi dovelo do toga da ovakvi faktori postanu zanemarljivi, čime bi se višestruko unapredila klinička praksa.

Ovo istraživanje fokusirano je na formiranje jedinstvenog sistema za detekciju patologije govora koji se bazira na aktuelnim dostignućima na polju obrade govora i standardnim metodama koje se u te svrhe koriste u logopedskoj praksi. U radu je prikazan razvoj sistema za procenu artikulacionih odstupanja govora koji je osmišljen tako da prati metodologiju logopedskog pristupa. Predstavljeni su rezultati modelovanja auditornog sistema čoveka prilikom detekcije akustičkih obeležja u govornom signalu koja učestvuju u manifestaciji patologije. Data su i istraživanja na temu segmentacije glasova pogodnih za dalju obradu radi utvrđivanja postojanja i stepena patologije. Prikazan je i rad na izradi automatskog sistema za klasifikaciju govora u smislu postojanja ili odsustva patologije u govornom signalu. Poseban deo posvećen je istraživanjima na temu prepoznavanja specifičnih artikulacionih odstupanja pojedinačnog glasa.

Rezultati istraživanja pokazali su da se upotrebom algoritama koji se koriste u obradi govora može izvesti sistem za automatsku procenu patološkog izgovora. Rezultati koji se dobijaju automatskom procenom u visokom procentu se podudaraju sa ocenama dobijenim od strane eksperata.

Ključne reči: obrada govornog signala, prepoznavanje patološkog govora, auditorni model, segmentacija govora.

Naučna oblast: elektrotehnika

Uža naučna oblast: akustika

UDK broj: 621.3

Title: Detection of articulatory-acoustic deviations in pathological speech

Summary

In clinical practice, dyslalia is usually assessed by subjective measures and methods, using the audio-visual evaluation performed by the trained experts - speech therapists. Qualitative and quantitative evaluation of pathological pronunciation is based on a number of standard tests and procedures. This approach could be influenced by factors arising from the examiner and the test conditions, which do not occur as a consequence of pathological pronunciation. The existence of an automatic system for the evaluation of speech pathology would make these factors negligible, and further improve clinical practice.

This research is focused on the system for speech pathology detection that is based on the current achievements in the field of speech processing and standard methods used for this purpose in speech therapy practice. In this dissertation, the development of the system for the assessment of articulation disorders is presented. The system is designed to follow the methods used in speech therapy. The model of human auditory system is presented, adjusted to the detection of the acoustic characteristics of the speech signal that participate in the manifestation of speech pathology. Also, research on automatic speech segmentation is described, which results in the segments suitable for further processing in order to determine the existence and degree of speech pathology. Subsequently, the system for classification of speech in terms of the existence or absence of speech pathology is introduced. A special section is devoted to the research on recognition of the specific articulatory deviations that occur during the pronunciation of individual phonemes.

The results showed that it is possible to develop the system for automatic evaluation of speech pathology using the speech processing algorithms. The results obtained by the automatic evaluation match to a great extent to the ones given by the experts.

Keywords: digital speech processing, speech pathology detection, auditory model, speech segmentation.

Scientific area: electrical engineering

Scientific subarea: acoustics

UDC number: 621.3

Sadržaj

Spisak slika

Spisak tabela

1 Uvod

Govorna komunikacija predstavlja najprirodniji oblik sporazumevanja između ljudi. Degradacija razumljivosti govora koja nastaje kao posledica patoloških pojava u artikulaciji govora utiče na kvalitet komunikacije, kao i kvalitet života u celini. Stručnjaci smatraju da sa terapijom govora treba započeti što ranije kako bi se povećale šanse za pravilan razvoj govora i jezika. Postupci dijagnoze i terapije podležu procedurama koje su usvojene u logopedskoj praksi. Evaluacija i reevaluacija kvaliteta artikulacije se vrše pomoću više testova i kliničkih postupaka na osnovu čega se usvajaju terapijski postupci koji odgovaraju određenom dijagnostifikovanom poremećaju ili grupi poremećaja, a u skladu sa mogućnostima konkretnog pacijenta. Tradicionalno, procenu vrše obučeni stručnjaci - logopedi, na osnovu adekvatnog obrazovanja i dugotrajnog usavršavanja u teoriji i praksi. I pored toga, odluka je do izvesne mere subjektivna. Takođe, procena kvaliteta govora izvršena od strane obučenih profesionalaca obično je vrlo zahtevna što se tiče vremena i ljudskih resursa što je dalje čini skupom i dugotrajnom i može dovesti do grešaka nastalih usled ljudskih faktora kao što su umor i nedostatak koncentracije. U cilju objektivizacije procene kvaliteta izgovora intenzivno se traga za rešenjima koja su neinvazivna, a dovoljno pouzdana i laka za korišćenje i koja su u saglasnosti sa ostalim elementima dijagnostike i terapije.

U poslednjih dvadesetak godina upotreba računarskih ekspertskih sistema baziranih na akustičkoj obradi govornog signala pokazala se kao dobra alternativa u smislu ispomoći kod terapije i dijagnostike poremećaja govora. Automatskim testiranjem uticaj ljudskog faktora je značajno smanjen, a rezultati su konzistentni. Takođe, čineći sistem dostupan putem Interneta ili javne telefonije, omogućeno je testiranje više ispitanika istovremeno, kako u specijalizovanim centrima koji se bave govornim poremećajima, tako i u kućnim uslovima, što mnogostruko smanjuje troškove

povećavajući istovremeno dostupnost, intenzitet tretmana i, konačno, vodi ka efikasnijoj habilitaciji i rehabilitaciji. Takav ekspertski sistem koji se odlikuje pristupačnošću, jednostavnošću korišćenja, efikasnošću detekcije patologije i koji je neinvazivan za pacijenta, može se koristiti i za masovne skrining testove, dijagnostiku i rano otkrivanje patologije govora. Pored toga, automatski sistem može biti od velike koristi i samim terapeurima, omogućavajući im da pomoću rezultata automatske detekcije patologije obrate pažnju na neke od patologija koje možda na drugi način ne bi uočili. Za pacijenta, on može da posluži kao alat za samoprocenu poremećaja govora i omogući mu da utvrdi koliko je ozbiljan poremećaj i treba li da zatraži pomoć specijaliste. Iako ovi sistemi ne mogu u potpunosti zameniti logopeda, oni mogu da daju procenu poremećaja govora i, kada je potrebno, upute ispitanika u specijalizovane medicinske ustanove na dalju analizu i tretman. Za logopeda, takav sistem bi mogao da se koristi komplementarno sa standardnim logopedskim metodama pružajući logopedu drugo mišljenje koje bi se moglo uzeti u obzir prilikom postavljanja dijagnoze. Takođe, takav sistem bio bi pogodan za praćenje i kontrolu toka tretmana pacijenta kao i kvalifikaciju promena kod govornih poremećaja. Osim toga, može se koristiti i u evaluaciji terapijskih metoda pokazujući koji metod daje bolje ukupne rezultate kada se testira na grupama pacijenata. Razvoj automatizovanog sistema zasnovanog na algoritmima obrade govora koji vrši detekciju patologije u govoru donosi mnoge prednosti na polju detekcije, terapije i korekcije patološkog govora.

Pouzdani sistem za automatsku detekciju govorne patologije pored toga što predstavlja potreban instrument kojim bi se unapredila logopedska praksa, takođe, predstavlja i izazov u krugovima istraživača koji se bave obradom govora. Ključni problem predstavlja definisanje onih akustičkih parametara pomoću kojih je moguće razlikovati oblasti normalnog i patološkog izgovora, kao i graničnih vrednosti tih parametara kojima se definiše razdvajanje navedenih oblasti. U tom smislu treba ispitati više vrsta parametrizacija govornog signala i utvrditi optimalni skup i međusobni odnos onih parametara sa kojima se postiže najbolja distinkcija. Poseban problem predstavlja činjenica da se detekcija i analiza patološkog govora standardno vrše audio-vizuelnim putem od strane stručnjaka sa treniranom percepcijom. To dalje znači da je potrebno izvesti određeni broj dobro osmišljenih perceptivnih eksperimenata, sve sa ciljem da se formira model percepcije patološkog govora.

1.1 Ciljevi istraživanja

Osnovni cilj ovog istraživanja jeste razvoj jedinstvenog modela za automatsko prepoznavanje patologije u govoru koji obuhvata metodološke aspekte koji se koriste u aktuelnoj logopedskoj praksi. Za realizaciju osnovnog cilja, predviđeni su sledeći podciljevi:

- formiranje jedinstvenog modela za prepoznavanje artikulaciono - akustičkih odstupanja glasova u patološkom izgovoru,
- postavljanje modela segmentacije glasova prilagođenom patološkom izgovoru,
- razvoj pojedinačnih modela za detekciju specifičnih oblika odstupanja akustičkih obeležja u vremenskom, amplitudskom, spektralnom ili parametarskom domenu,
- komparativna analiza više algoritama za prepoznavanje i klasifikaciju odstupanja u artikulaciji glasova prema kliničkim kriterijumima.

1.2 Metodologija istraživanja

Namera je da se na osnovu sistematizacije istraživanja koja su prisutna u literaturi a bave se temom percepcije patološkog govora i matematičkih algoritama koji se tiču obrade govora, formira algoritam za prepoznavanje artikulaciono-akustičkih patologija u govoru. Ovakav algoritam mogao bi biti implementiran i kao sastavni deo nekog šireg sistema koji obuhvata sve elemente logopedske prakse.

U tom cilju, metodologija se zasnivala na nekoliko pristupa. Osnovu predstavlja modelovanje logopedskog pristupa u percepciji i oceni patologije i stepena patologije u glasu. Sa tim u vezi potrebno je formirati bazu patološkog i normalnog izgovora i adekvatno je obraditi. Obrada je podrazumevala ocenjivanje od strane eksperata i segmentaciju na manje fonemske jedinice koje su dalje analizirane. Baza je služila kako u svrhu istraživanja akustičkih karakteristika specifičnih za prepoznavanje odstupanja na nivou određenog glasa tako i u definisanju algoritama za obradu govornog signala i

detekciju ovih obeležja. Takođe, baza je korišćena i za obuku i testiranje predloženih algoritama. Prateći logopedski pristup detekciji patologije u govoru, javljaju se dva nivoa procene poremećaja artikulacije. Prvi je predstavljan algoritmom za globalnu procenu postojanja patologije i predstavlja kategorizaciju po stepenu patologije. Drugi nivo je detaljnija analitička ocena odstupanja na nivou grupe obeležja karakterističnih za dati glas.

Razvoj softvera za analizu i detekciju patologije, počevši od parametrizacije govornog signala, preko segmentacije do detekcije obeležja, je izveden na osnovu aktuelnih trendova u računarskoj analizi patološkog govora, kao što su modeli artikulaciono - akustičkog odstupanja u govoru, auditorni modeli, neuralne mreže i drugo. Takođe, korišćeni su raspoloživi softverski alati za analizu govornih signala.

Dve osnovne karakteristike ovog sistema, a koje su imale presudan uticaj na izbor parametara, su činjenica da je govorni signal koji se analizira unapred poznat, kao i činjenica da algoritam mora biti robustan sa obzirom na to da se koristi u svrhu detekcije različitih oblika patološkog govora.

Osnovni koraci u dobijanju eksperimentalnih rezultata su**:**

- Analiza dosadašnjih istraživanja na polju detekcije patologije u govoru.
- Izbor govornog korpusa.
- Predobrada govorne baze, odnosno segmentacija i ocenjivanje stimulusa od strane eksperata.
- Segmentacija govornog korpusa predloženim algoritmima, statistička obrada podataka i analiza dobijenih rezultata.
- Poređenje algoritama za globalnu ocenu patologije govora i analiza rezultata.
- Komparativna analiza više algoritama za prepoznavanje i klasifikaciju odstupanja u artikulaciji glasova prema kliničkim kriterijumima.

1.3 Kratak opis sadržaja disertacije

Istraživanja su sistematizovana u okviru poglavlja koja slede.

U poglavlju 2 data je definicija pojma patologije govora a navedene su i tipične manifestacije patološkog izgovora. Opisani su osnovni principi u okviru standardnog postupka evaluacije artikulacionih poremećaja. Dat je prikaz dva logopedska testa koja se u tu svrhu koriste u kliničkoj praksi Instituta za eksperimentalnu fonetiku i patologiju govora u Beogradu za procenu artikulacionih poremećaja govora.

Poglavlje 3 posvećeno je sistematizaciji dosadašnjih istraživanja na temu detekcije patologije u govoru. Tu su prikazani i postojeći automatski (računarski) sistemi koji su trenutno u različitim fazama razvoja a koji predstavljaju pomoć u terapiji i koji u određenoj meri podržavaju neku od vrsta detekcije poremećaja izgovora.

U poglavlju 4 predložen je sistem za evaluaciju poremećaja artikulacije za srpski jezik. Pored opisa celog sistema, prikazani su i njegovi moduli: modul za predobradu koji u sebi sadrži deo za izdvajanje obeležja i segmentaciju govornog signala, modul globalnog artikulacionog testa kao i modul analitičkog testa. Pored toga, date su i karakteristike baze stimulusa koja se sastoji od uzoraka patološkog i normalnog izgovora, a koja je korišćena u okviru istraživanja sprovedenih u ovoj disertaciji.

Istraživanja vezana za segmentaciju govornog signala, kako normalnog tako i patološkog govora, i različiti pristupi tom problemu dati su u poglavlju 5. Prikazana su dva pristupa segmetaciji govora i to segmetacija pomoću algoritma maksimalne verodostojnosti (*Maximum Likelihood* - ML) i segmentacija pomoću dinamičkog vremenskog usklađivanja (*Dynamic Time Warping* – DTW). Pored standardnog algoritma maksimalne verodostojnosti razmatran je i njegov modifikovani oblik koji uključuje isticanje naglih promena u spektru govornog signala. Kod algoritma vremenskog usklađivanja korišćene su dve grupe ulaznih parametara, Mel frekvencijski kepstralni parametri (*Mel frequency cepstral coefficients* - MFCC) i Gama frekvencijski kepstralni parametri (*Gamma frequency cepstral coefficients* - GFCC).

U poglavlju 6 su sabrana sprovedena istraživanja vezana za automatsku detekciju patologije govora na bazi globalne ocene kvaliteta artikulacije. Prikazan je pristup u kom su korišćene neuralne mreže, dat je kratak pregled iz literature vezan za detekciju patološkog izgovora korišćenjem ove metode. Opisan je postupak pronalaženja adekvatnog rešenja za ovaj kompleksan problem koji je podrazumevao usložnjavanje topologije neuralne mreže i adekvatan izbor obeležja govornog signala.

Prikazani su rezultati istraživanja koji su dobijeni korišćenjem višeslojnog perceptrona i ansambla neuralnih mreža.

Deo sistema koji se bavi automatskom detekcijom pojedinačnih artikulacionih poremećaja na bazi analitičke ocene dat je u poglavlju 7. U okviru ovog poglavlja posmatrana su dva obeležja patološkog glasa: stridens i poremećaj trajanja glasa. Prikazane su artikulaciono-akustičke karakteristike ovih patologija izgovora kao i teorijska osnova na kojoj su bazirani algoritmi za automatsku detekciju. Dati su rezultati istraživanja na polju razlika tipičnog i atipičnog izgovora koji su služili kao osnova automatskog pristupa. Detaljno je prikazan automatski postupak detekcije stridensa baziran na auditornom modelu i njegovo poređenje sa postupkom koji ne uključuje psihoakustičke efekte. Pored algoritma za detekciju stridensa, predstavljen je i algoritam za detekciju poremećaja trajanja glasa, koji u osnovi ima algoritam za segmentaciju. Na kraju, prikazani su rezultati primene ovih algoritama za detekcije artikulacionih poremećaja.

U poslednjem, osmom poglavlju, kroz pregled rezultata opisanih algoritama i metoda primenjenih na uzorke normalnog i patološkog izgovora, predstavljen je ukupan uvid u mogućnosti predloženog sistema za automatsku detekciju patološkog izgovora. Na osnovu toga, istaknut je niz konkretnih rezultata koji predstavljaju najznačajniji doprinos ove disertacije. Zahvaljujući modularnom pristupu koji je primenjen pri rešavanju problema automatske detekcije patologije govora, ostavljen je prostor za dalju nadogradnju i razvoj što je i diskutovano na samom kraju ovog poglavlja kroz mogućnosti daljih istraživanja.

2 Patologija govora

Patologija govora i jezika može se manifestovati na različitim nivoima govorno-jezičkog razvoja. Razlozi za pojavu nepravilnog izgovora glasova, bilo da se javljaju prilikom izgovora izolovane foneme ili u širem govornom kontekstu, mogu biti fiziološke, psihološke ili kognitivne prirode. Ove nepravilnosti se manifestuju u vidu odstupanja u izgovoru u odnosu na izgovorne norme koje odgovaraju zadatom jeziku. U logopedskoj praksi navedene norme razvoja govora i jezika prate psihološki i fiziološki razvoj. Nakon razvojnog perioda smatra se da je govor automatizovan i podrazumeva se pravilna artikulacija. Međutim, ukoliko se atipičan izgovor i dalje zadrži nakon perioda predviđenog razvojnim normama to znači da nepravilnosti u izgovoru nisu razvojne prirode i takav izgovor se smatra patološkim. Tipovi devijacija u izgovoru variraju u zavisnosti od glasovne grupe jednog jezika, kao i između različitih jezika za istu fonemu.

Dislalija je jedan od oblika patologije govora i jezika i ispoljava se kao atipična produkcija glasova, odnosno patološka artikulacija koja za posledicu ima atipični glas u akustičkom domenu. Nastaje kao posledica dislokacije, odnosno, nepravilnog položaja jednog ili više organa koji učestvuju u artikulaciji, što za posledicu ima nepravilan izgovor fonema. Dislalija je najčešći poremećaj govora kod dece i može se manifestovati kao: omisija, odnosno nedostatak nekih glasova; supstitucija, tj. zamena nerazvijenog glasa glasom koji već postoji; i distorzija koja predstavlja različita tipična i atipična oštećenja pojedinih izgovornih glasova.

Postoje specifični tipovi odstupanja koji su karakteristični za svaku od grupa glasova. Ova odstupanja su određena samom strukturom glasova i glasovnim atributima koji čine određeni fonem. Neka od odstupanja glasova, u odnosu na pripadnost glasovnoj grupi, prikazana su i opisana u istraživanjima Punišić i sar. (2007, 2011a,b, 2012). U tipične distorzije izgovornih glasova spadaju: interdentalni sigmatizam

(vrskanje), lateralni sigmatizam (šuškanje) i nazalni sigmatizam - snorting (unjkav izgovor) (Vladisavljević, 1981; Golubović,1997).

Frekventnost poremećaja artikulacije nije ista za sve glasove. U srpskom jeziku, odstupanja se najčešće javljaju kod konsonanata, dok su odstupanja u izgovoru vokala veoma retka. U srpskom jeziku se izdvaja grupa od 12 kritičnih glasova koji najčešće odstupaju od tipičnog izgovora, kod dece ali i kod odraslih govornika, sa frekvencijom odstupanja (u %): /s/ - 12,6; /c/ - 11,5; /z/ - 11,4; /č/ - 9,8; /š/ - 9,6; /ž/ - 8,8; /dž/ - 8,1; /r/ - 8,0; /ć/ - 5,1; /l/ - 3,7; /lj/ - 3,5; /đ/ - 0,6 (Jovičić, 1999). Svi ovi fonemi sadrže izrazito šumni deo spektra u visokofrekvencijskom području i zahtevaju precizno pozicioniranje artikulacionih organa kod generisanja fonema.

Prepoznavanje elemenata patologije prilikom produkcije jedne foneme je složen perceptivni proces. U logopedskoj praksi, zahvaljujući dobro razvijenom naučnom pristupu patologiji govora, ustanovljena je metodologija kojom je omogućena ocena stepena poremećaja na različitim nivoima artikulacije (Kostić i sar., 1983). U analizi akustičkih karakteristika govora u tradicionalnoj logopedskoj dijagnostici, tretmanu i evaluaciji primenjuje se ekspertsko (trenirano) slušanje. Njime se postojanje glasovnih odstupanja ocenjuje auditivno perceptivnom procenom, odnosno sluhom. Procenu daju obučeni eksperti - logopedi na osnovu iskustva u oblasti standardnih artikulacionih i akustičkih karakteristika glasova. Metodologija ispitivanja je takva da se od ispitanika očekuje da produkuje govorni signal, dok ispitivač sluša i analizira. U zavisnosti od uzrasta ispitanika i svrhe procene izgovora govorni materijal mogu biti izolovane reči, rečenice ili tekst, a ponavljanje se može vršiti samostalnim čitanjem ili ponavljanjem stimulusa za ispitivačem. Analizirajući karakteristike izgovorene foneme beleži se postojanje ili odsustvo unapred definisanih katakeristika koje utiču na nepravilnost izgovora i njihov uticaj na stepen atipičnosti. Za procenu kvaliteta izgovornih glasova u srpskom jeziku se, između ostalog, koriste Globalni artikulacioni test (GAT) i test za Analitičku ocenu artikulacije srpskog jezika (AT). Ovi testovi se koriste u kliničkoj praksi Instituta za eksperimentalnu fonetiku i patologiju govora u Beogradu (Kostić i sar., 1983; Punišić, 2012).

2.1 Globalni artikulacioni test

Globalni artikulacioni test (GAT) (Kostić i sar., 1983) se upotrebljava kod procene kvaliteta i nivoa dostignute razvijenosti izgovora glasova srpskog jezika. U testu se ocena daje na osnovu ukupne audiovizuelne predstave o izgovornom glasu. Test čini 30 reči, gde svaka reč sadrži glas od interesa za analizu. Ispitivani glasovi nalaze se u inicijalnoj poziciji u slučaju konsonanata, odnosno u interkonsonantskoj poziciji kod ispitivanja vokala. Govorni stimulusi artikulacionog testa dati su u tabeli 2.1. Kvalitet izgovora vrednuje se ocenama od 1 do 7. Glasovi izgovoreni u skladu sa standardnom normom za srpski jezik označavaju se ocenama 1, 2 ili 3 i smatraju se tipičnim. Glasovi koji odstupaju od tipičnog izgovora odgovaraju ocenama 4, 5 ili 6 i ukazuju na odstupanje po tipu distorzije. Težina distorzije srazmerna je visini ocene. Ocena 4 ukazuje namanje odstupanje od pravilnog izgovora, a ocena 5 oštećenje po mestu i/ili načinu tvorbe. Ocenu 6 dobijaju izrazito distorzovani glasovi koji su toliko oštećeni da se van konteksta ne mogu prepoznati. Ocenu 7 dobijaju omitovani ili supstituisani (zamenjeni drugim) glasovi. Svaka ocena je proizvod svih artikulaciono - akustičkih obeležja koja doprinose ukupnom utisku o kvalitetu izgovornog glasa.

Tabela 2.1 Reči Globalnog artikulacionog testa (GAT).

Reči					
i-vidi	p-pada	g-guma	dž-džep	š-šuma	m-moj
e-beba	b-baba	c-cica	f-fes	ž-žaba	n-noga
a-mama	t-tata	ć-ćebe	v-voz	h-hodi	nj-njiva
o-voda	d-deda	đ-đak	s-seka	j-jaje	l-lice
u-buba	k-koka	č-čelo	z-zima	r-riba	lj-ljudi

2.2 Analitički test

Test za Analitičku ocenu artikulacije (AT), uz pomoć detaljne analize svih elemenata koji utiču na nepravilnu realizaciju foneme u akustičkom domenu,

omogućava identifikaciju (određivanje) vrste distorzije, supstitucije ili omisije. Analitički test se izvodi imenovanjem pojmova koji se nalaze na slikama, njih 100. Tom prilikom, ispitanik izgovara posmatrani glas u inicijalnom, medijalnom i finalnom položaju. Svakom od 5 vokala odgovara po pet reči, a za preostalih 25 glasova po tri reči. Svi glasovi su podeljeni po glasovnim grupama. Test sadrži i 30 rečenica pomoću kojih se proverava spontana upotreba glasova. Za svaku glasovnu grupu definisan je skup obeležja po kojima izgovorni glas može odstupati od svoje tipične realizacije. Moguća odstupanja određenog glasa zavise od glasovne grupe kojoj pripada i data su u vidu: trajanja glasa, visine osnovnog tona, osobina zvučnosti, stepena nazalnosti, područja i načina artikulacije, itd. Postoje tačno definisana obeležja, po kojima jedan glas može odstupati koja se mogu odnositi na cele grupe glasova ili su potpuno specifična i odnose se na tačno određeni glas. Na taj način dobija se informacija o stanju svih elemenata artikulacije jednog ispitanika i stiče se uvid u najfrekventnija odstupanja i nedostatke u izgovoru na osnovu kojih se mogu odrediti najadekvatnije mere habilitacije i rehabilitacije. AT se koristi za: dijagnostiku vrste i stepena artikulacionog odstupanja; planiranje korekcionog postupka; precizna naučna istraživanja u oblasti patologije govora, a može se koristiti i za proučavanje razvoja govora.

Tabela 2.2 Glasovna odstupanja kod frikativa

1. bezvučno	11. bilabijalno V	21. alveolarizovano
2. zvučno	12. stridens	22. palatalizovano
3. produženo	13. koronalno	23. preoštro H
4. skraćeno	14. interdentalno I stepen	24. H pomereno nazad
5. jaka frikcija	15. interdentalno II stepen	25. guturalno H
6. slaba frikcija	16. interdentalno III stepen	26. pregradno H
7. visoko	17. adentalno	27. unilateralno
8. nisko	18. zaokružene usne za S	28. neodređeno
9. nazalizovano	19. desna lateralna frikcija	29. nema glasa
10. bilabijalno F	20. leva lateralna frikcija	30. centralni glas
	31. substitucija	

U ovom radu, AT je modifikovan u tom smislu da su stimulusi bile reči iz artikulacionog testa dok su ocene po svemu odgovarale opisanom analitičkom testu. Za ilustraciju karakteristika koje se utvrđuju artikulacionim testom u tabeli 2.2 dat je prikaz analiziranih glasovnih odstupanja za frikative. Sličan spisak odstupanja postoji za svaku glasovnu grupu (Kostić i sar., 1983).

3 Pregled literature - dosadašnja istraživanja

Sa razvojem dostignuća na polju obrade akustičkog signala, u prvom redu sistema za automatsko prepoznavanje govora i različitih metoda za parametrizaciju i klasifikaciju govora, istraživači su prepoznali potencijal primene ovih metoda za klasifikaciju i prepoznavanje patološkog govora kao i detekciju patoloških stanja u govoru.

Počevši od 70-tih godina prošlog veka, raslo je interesovanje istraživača za pojave vezane za neregularne promene u akustičkim karakteristikama govora. Fujimura (Fujimura i sar., 1971) detaljno je analizirao varijacije u govornom signalu nastale kao posledica različitih načina artikulacije prilikom izgovora različitih fonema. Nakon toga, razni autori posvetili su se istraživanjima vezanim za analizu i prepoznavanje anomalija u izgovoru koje se manifestuju u akustičkom domenu.

Najčešće razmatrani problemi su bili vezani za detekciju i analizu disfonije, polipa, vokalnih nodula (čvorića na glasnicama) , unilateralne laringealne paralize, kancera glotisa i drugih laringealnih oboljenja. U jednom broju radova obrađivane su teme vezane za kvalitet izgovora kod osoba sa Parkinsonovom bolešću, dizartrijom, afazijom kao i artikulacionim poremećajima. Glavni napori istraživača bili su orijentisani na proučavanje akustičkih parametara i tehnika klasifikacije kako bi se postigla velika tačnost razlikovanja između normalnog i patološkog izgovornog glasa. Generalno gledajući, postoje dve grupe rešenja. Prvoj pripadaju postupci detekcije patologije u govoru koji su fokusirani na primenu i razvoj novih parametara za meru kvaliteta govora, dok drugoj pripadaju oni koncentrisani na razvoj boljih računarskih metoda i algoritama za automatsku diskriminaciju između normalnog i patološkog govora.

Pored ovoga, postoje softverski sistemi koji se koriste kao ispomoć u logopedskoj terapiji, u okviru kojih se vrši detekcija poremećaja govora, kao i neki od programskih procedura za detekciju artikulacionih poremećaja.

3.1 Karakteristike istraživanja

Najveći broj parametra koji se može naći u literaturi u domenu automatske detekcije patologije u govoru bazira se na dugovremenoj analizi signala (Boyanov i sar., 1997; de Krom, 1993; Feijoo i sar., 1990; Kasuya i sar., 1986; Michaelis i sar., 1997; Manfredi, 2000; Hillman i sar., 1997; Yumoto i sar., 1982). Ovi dugovremeni parametri se računaju usrednjavanjem lokalnih vrednosti promena amplituda (*shimmer*), frekvencija (*jitter*) (Feijoo i sar., 2000) i/ili estimacije šuma. Pokazalo se da parametri koji se tiču šuma mogu biti vrlo pouzdani indikatori poremećaja glasa jer patologija unosi određenu dozu šuma u glas. Ova ispitivanja su uglavnom rađena na kontinualno izgovaranim vokalima. Najčešće korišćeni parametri su SNR (Klingholtz i sar., 1987), *Harmonic to Noice Ratio* (HNR) (Yumoto i sar., 1984; de Krom, 1993), *Normalized Noice Energy* (NNE) (Kasuya i sar., 1986), *Voice Turbulence Index* (VTI) (Yumoto i sar., 1982), i *Glotal to Noice Excitation Ratio* (GNE) (Michaelis i sar., 1997). Gordino i sar. (2010) izvršili su poređenje efikasnosti GNE, NNE, CHNR i HNR u detekciji patološkog glasa i dobili slične rezultate za prva tri parametra, među kojima GNE ima prednost jer ne zahteva računanje *pitch* periode.

Još jedna grupa obeležja koja se standardno koriste su MFCC koeficijenti. Dibazar i sar. (2002) koristili su kombinaciju *pitch* periode i MFCC sa HMM (*Hidden Markov Model*) kao klasifikatorom na 710 patoloških i 53 normalna izgovora kontinualno izgovanog glasa /a/ i dobili rezultate od 98,59% prepoznavanja patološkog glasa.

Douglas Cairns i sar. (1994) predložili su tehniku za detekciju hipernazalnosti pomoću *Teager Energy Operatora* (TEO) obeležja (Kaiser, 1990). Koristeći HMM za klasifikaciju normalnog i hipernazalnog izgovorana 11 zdravih i 11 simuliranih hipernazalnih uzoraka, dobijeno je prepoznavanje od 100% za hipernazalne uzorke i 98,8 za zdrave uzorke. Cairns i sar. (1996) su prikazali rezultate unapređenog algoritma

koji se bazira na TEO obeležju i koristi *likelihood detector* u fazi klasifikacije. Skorije studije (Hadjitodorov i sar., 2002) koriste nekoliko već pomenutih obeležja kao i *turbulent noice estimation* obeležje za automatsku detekciju patologije glasa.U ovom slučaju, sistem je pokazao tačnost od 96.1% koristeći kNN (K-*nearest neighbors*) i komercijalno dostupnu bazu MEEI (Massachusetts Eye and Ear Infirmary, 1994.). Uzorci su bili u obliku kontinualno izgovaranog glasa /a/ od strane 691 osobe, od čega je 638 osoba imalo neki od poremećaja (funkcionalnih ili organskih) laringealnog glasa dok je 51 osoba bila bez poremećaja glasa.

Maier i sar. (Maier i sar., 2008; Maier i sar., 2006) su posmatrali razne prozodijske karakteristike govora kod automatske detekcije hipernazalnosti kod dece sa CLP (*Cleft Lip and cleft Palate*). Kombinacijom ovih karakteristika i 24 MFCC koeficijenta, postignuto je prepoznavanje hipernazalnosti od 71,1% na nivou vokala i 75,8% na nivou reči na uzorku od 26-oro dece sa CLP.

Suočavajući se sa istim problemom tj, sa hipernazalnošću, Murillo i sar. (Murillo i sar., 2011.) su prikazali kako mere šuma kao što su HNR, CHNR, NNE i GNE daju relevantne informacije kod detekcije hipernazalnosti kod dece sa CLP. Pored navedenih mera, autori su implementirali još i *jitter*, *shimmer* i 11 MFCC koeficijenata. Korišćenjem PCA (*Principal Commponent Analysis*) izabrali su najznačajnije mere za detekciju hipernazalnosti. Baza se sastojala od 1280 snimaka 5 kontinualno izgovaranih vokala španskog jezika (256 po vokalu od kojih 110 normalnih, a 146 hipernazalnih). Upotrebljen je linearni klasifikator i dobijena je tačnost prepoznavanja od 78,86% do 88.82% u zavisnosti od vokala. Autori su pokazali da mere šuma mogu i do 20% poboljšati klasifikaciju.

Delgado i sar. (Delgado i sar., 2011) prikazali su metod detekcije hipernazalnosti posmatrajući plozive. Baza se sastojala od izgovora 88-oro dece, 44 normalnih i 44 patoloških stimulusa (reči /koko/ i/papa/). Rezultati su pokazali tačnost prepoznavanja od 85,2% za /k/ i 89.5 za /p/. Kada su posmatrane obe foneme, tačnost je bila 92,7%. Ovo je važna studija jer se hipernazalnost obično detektuje na zvučnim fonemama.

Titze i sar. (1993, 1995) pošli su od činjenice da su vibracije glasnica nelinearni fenomen i klasifikaciju glasova su tumačili prema nelinearnoj dinamici (*Non Linear Dynamics* - NLD). Nakon demonstracije nelinearne dinamike kod produkcije, mnogi

autori su se fokusirali na karakterizaciju govornog signala koristeći nelinearne metode (Matassini i sar., 2000; Alonso i sar., 2001, 2005; Zhang i sar., 2004, 2005). Goivanni i sar. (1999) koristili su *Largest Lyapunov Exponent* (LLE) kao alternativu procesu automatske detekcije patologije. Korišćeni su uzorci koji potiču od 12 zdravih osoba i 26 osoba sa unilateralnim laringealnom paralizom. Baza uzoraka sastojala se od kontinualno izgovaranog glasa /a/. Rezultati su pokazali da normalan izgovor ima vrednost LLE od 0.38 ±0.182 dok patološki ima vrednost 0.57±0.337 i daje ova karakteristika korisna za analizu patološkog glasa. Najpoznatiji pristup na polju nelinearne dinamike dat je u (Jiang i sar., 2006) gde se koriste nelinearne karakteristike kao što su *Lyapunov exponents* i *Kolmogorov entropy*. Autori daju zaključak da se kombinacijom tradicionalih parametara i nelinearne dinamičke analize mogu poboljšati rezultati analize patološkog i laringealnog glasa.

Poslednjih godina, uspešno se koriste višedimenzionalni parametri koji se sastoje od kombinacija akustičkih obeležja govora kao što su MFCC, NNE, GNE i HNR sa nelinearnim dinamičkim parametrima (Little i sar., 2011) postižući rezulatate prepoznavanja preko 98% (Arias i sar., 2011).

U cilju klasifikacije patološkog govora u literaturi se mogu naći različiti pristupi od jednostavnih kao što su kNN i *Linear Discriminant Analysis* (LDA) (Shama i sar., 2007), do složenijih kao što su HMM i *Support Vector Machines* (SVM). HMM (Dibazar i sar., 2002, 2006) daje dobre rezultate (od 73% do 98%) u kombinaciji sa MFCC koeficijentima i različitim merama šuma. SVM (Gelzinis i sar., 2008) pokazuje najbolje rezultate (od 92% do 95%) kada se koristi sa *wavelet* dekompozicijom, linearnim prediktivnim koeficijentima (Silva-Fonseca i sar., 2007), merama šuma i MFCC koeficijentima i njihovim izvodima (Saenz-Lechon i sar., 2008). Pored navedenih, uspešno se koriste i ANN (*Artificial Neural Network*) (Ritchings i sar., 2002; Linder i sar., 2008; Fraile i sar., 2009) i statistički pristupi bazirani na *Gaussian Mixture Model* - GMM (Godino-Llorente i sar., 2006).

3.2 Rešenja namenjena terapiji govora

Počevši od 90-tih godina prošlog veka, javlja se interesovanje istraživača za razvoj softverskih rešenja koja bi mogla biti upotrebljena u okviru terapije govora i

jezika. Kako je sama terapija govora i jezika obiman i komplikovan proces, razvijena su rešenja koja delimično ili u potpunosti predstavljaju pomoć obučenim terapeutima. Karakteristike i konkretna namena tih rešenja su raznolike. Neka od njih su osmišljena da u najvećoj meri pruže pomoć kod praćenja i organizacije logopedskog tretmana u smislu inteligentnih sistema kojima se analiziraju tok lečenja i podataka generisanih od strane standardizvanih testova i daju sugestije terapeutima o daljem toku terapije. Druga su, opet, osmišljena kao integrisani moduli za automatsku procenu poremećaja govora i jezika, ili predstavljaju module za vežbe govora i jezika. Neka od njih su dizajnirana za kućnu upotrebu i uporebu u kliničkoj praksi, međutim ima i onih koja su prilagođena za prenosne uređaje ili dostupna putem Interneta. U nastavku će biti navedene karakteristike nekih od njih koji obuhvataju analizu i detekciju poremećaja govora, kao i neka softverska rešenja koja se bave automatskom detekcijom artikulacionih poremećaja.

CATSEAR (Turk i Arslan, 2005) je softver dizajniran za prikupljanje govornih signala, analizu podataka, dizajn i praćenje govorne terapije. Primena je ograničena na korekciju umerenih problema kod izgovora i govorni trening. Automatska procena kvaliteta govora koja se vrši pomoću tehnika prepoznavanja obrazaca (oblika) omogućava logopedima da upotrebe objektivne kriterijume tokom terapije govora kao i da pomognu pacijentima u situacijama kada terapeut nije prisutan. *Catsear*-om se automatski vrši procena kvaliteta izgovora na nivou foneme, reči i rečenice kako bi dao što detaljniji opis kvaliteta artikulacije. Modul kojim se procena vrši bazira se na upotrebi HMM-a na obeležjima kao što su *jitter, shimmer* i odnos energije visokih i niskih harmonika.

Softverski paket **Vocaliza** (Vaquero i sar.,2006, 2008) razvijen je kao deo projekta za izradu polu-automatskog sistema za terapiju govora i jezika ‘‘Comunica” (Saz i sar., 2009; Escartin i sar., 2008) u okviru istraživanja u oblasti govornih tehnologija na institutu Aragon, Univerzitet u Saragosi, Španija. *Vocaliza* je aplikacija orijentisana na govorni trening artikulacionih sposobnosti pacijenata dečijeg uzrasta na izolovanim rečima i kratkim rečenicama. Uz pomoć slika i animacija, deca se stimulišu da izgovaraju određene govorne sadržaje, unapred specificirane od strane terapeuta kako bi se fokusirali na konkretne potrebe svakog pojedinačnog korisnika. Kao rezultat dobija se ocena kvaliteta izgovora. *Vocaliza* se oslanja na više elemenata govornih

tehnologija koji čine sistem za prepoznavanje govora (*Automatic Speech Recognition* - ASR), sistemu za sintezu govora, sistem za prilagođavanje govorniku i sistem za procenu izgovora (*Pronutiation Verification* - PV). Centralni deo je svakako ASR koji se koristi za robustno prepoznavanje govora kako bi sistem utvrdio koje su reči izgovorene. On je izveden pomoću HMM i GMM sa vektorom obeležja koga čine 12 MFCC koeficijenata, energija, i njihovi prvi i drugi izvodi. Sinteza govora omogućava formiranje pravilnog izgovora što je značajno za uspostavljanje modela izgovora prilikom terapijskog postupka. Modul za procenu kvaliteta izgovora omogućava evaluaciju na nivou reči pomoću *Likelihood Ratio* (LR) *Utterance Verification* (UV) procedure (Lleida i Rose, 2000) kako bi se dodelila mera pouzdanosti izgovorene reči. Testiranja izvedena na različitim govornim bazama pokazuju rezultate koji su vrlo blizu onima dobijem od strane eksperata.

OLP (*Orto-Logo-Paedia*) projekat (Oster i sar., 2002) je projekat nastao u okviru saradnje Instituta za jezik i obradu govora u Atini sa sedam drugih partnera iz akademskog i medicinskog domena. OLP se sastoji od tri modula: OPTACIA, GRIFOS i Telemachos, i predstavlja softverski alat predviđen za korišćenje kao dopuna u terapiji govora za specifične poremećaje artikulacije koji integriše automatsko prepoznavanje govora i učenje na daljinu. OPTACIA (*opto-acoustic articulography*) modul ima za cilj da u realnom vremenu klijentu pruži vizuelne infromacije o položaju i pokretima artikulacionih organa korišćenjem 2D ili 3D prikaza vokalnog trakta. Na taj način se daje primer ispravnog položaja artikulacionih organa i pomaže kod uspostavljanja pravilne artikulacije. GRIFOS predstavlja sistem za prepoznavanje govora malog rečnika koji je zavistan od govornika. Služi za kvantitativnu analizu izgovora na nivou sloga i reči tokom terapije. Takođe, može se koristiti i za procenu artikulacije kontinualnog govora. Telemachos je deo odgovoran za učenje na daljinu i udaljeni monitoring. On omogućava proširenje i nastavak terapije i van kliničke institutcije i njenu primenu u kućnim uslovima, smanjujući na taj način troškove terapije kao i efikasnost terapije i evaluaciju učinka na daljinu.

STAR (*Speech Training Assessment Remediation*) sistem (Bunnell i sar., 2000a,b) dizajniran je kao pomoćno sredstvo za govorno-jezičku terapiju prilikom tretmana dece sa artikulacionim poremećajima. On je zajednički projekat "Alfred I. duPont" dečije bolnice iz Vilmingtona (savezna država Delaver, Sjedinjene Američke

Države) i Univerziteta u Delaveru (Sjedinjene Američke Države). Sistem je namenjen da terapeutima (logopedima) pruži pomoć prilikom procene kvaliteta izgovora omogućavajući im da postave ciljeve terapeutske intervencije, povećaju efikasnost terapije repetativnim vežbama artikulacije i pomognu kod prikupljanja i čuvanja podataka i pisanja izveštaja. Koristeći animirani karakter u okviru kompjuterske igre, konstantno se evocira govor kod deteta i ocenjuje kvalitet izgovora. Sistem je namenjen postepenom treningu najpre izolovanih glasova, zatim glasova u okviru reči i na kraju rečenica. Sistem je testiran na artikulaciji fonema kod dece uzrasta od 4 do 6 godina koja su izgovarala unapred zadate izolovane reči. Inicijalni fonem je bio ocenjen od strane eksperata. Sistem se bazira na diskretnom HMM, treniranom na osnovu govora dece sa normalnom artikulacijom. Rezultati su pokazali da pravilno treniran DHMM (*Descrete Hidden Markov Model*) može da obezbedi tačnu klasifikaciju (Bunnell i sar., 2000a).

PEAKS (*Program for Evaluation and Analysis of all Kind of Speech*) je okruženje za snimanje i analizu namenjeno automatskoj ili manualnoj evaluaciji govora i govornih poremećaja (Maier i sar., 2008, 2009.). Razvijen je na univerzitetu Erlangen-Nurnberg, Erlangen u Nemačkoj i koristi se na Univerzitetskoj klinici u naučne svrhe. Govor se analizira pomoću modula za automatsko prepoznavanje govora (ASR) i prozodijskog modula. ASR modul je zasnovan na skrivenim Markovljevim modelima (HMM) a govor je parametrizovan sa 11 MFCC koeficijenata, energijom signala kao i njihovim prvim izvodima, dok je u prozodijskom modulu izdvojeno 21 obeležje u govornom signalu. Sistemu se može pristupiti putem Interneta ili telefona i ne zahteva nikakav specijalni hardver, osim standardnog računara i zvučne karte. Bazira se na standardizovanom testu za nemački jezik u kome ispitanik čita tekst ili imenuje slike. Evaluacija kvaliteta izgovora vrši se na skali od 0 do 5. Sistem je testiran na dve grupe poremećaja: poremećaje glasa i artikulacione poremećaje. Testiranje je sprovedeno na ispitanicima kojima je izvršena rekonstrukcija larinksa nakon laringektomije (uklanjanje larinksa) - 41 ispitanik i dece sa rascepom usne i nepca -31 ispitanik. U prvom slučaju, kvalitet glasa je niži u poređenju sa normalnim govorom, odnosno, nakon rekonstrukcije larinksa javlja se "hrapav" glas, a dinamičke karakteristike govora kao što je *pitch* su ograničene što dovodi do monotonog govora. Kod dece sa rascepom usne i nepca karakteristike govornih odstupanja su uglavnom u formi raznih artikulacionih

poremećaja, tj. nazalizovan govor, promene u položaju artikulacionih organa koje dovode do substitucije (npr. glasa /d/ glasom /g/) kao i slabljenje tenzije artikulatora (npr. slabljenje ploziva). Rezultati su pokazali značajnu korelisanost između automatske ocene dobijene *leave one out* metodom i onih dobijenih ekspertskom analizom. Dobijena je korelisanost od 0.9 za laringektomiju i 0.87 za artikulacione poremećaje što je u granicama međuekspertske greške (Maier i sar, 2009).

TERAPERS (Danubianui sar., 2009) razvijen u Centru za kompjutersko istraživanje Univerziteta "Stefan cel Mare" iz Suceava, Rumunija, je inteligentni sistem dizajniran za pomoć kod trapije dislalije za decu predškolskog uzrasta i optimizovan je za rumunski jezik. Cilj ovog projekta je razvoj ekspertskog sistema za personalizovanu terapiju govornih poremećaja koji omogućava projektovanje toka govorne terapije, prilagođavanje terapije kategoriji poremećaja govora, procenu napretka i razvoj terapijskog materijala koji omogućava kombinovanje klasičnih metoda sa pomoćnim postupcima baziranim na audio-vizuelnom materijalu. Sistem se sastoji 4 modula i to: (1) programa za praćenje terapije, (2) 3D artikulacionog modela (koji prikazuje tačan položaj obraza, usana, zuba i jezika kod produkcije govora), (3) modula za kućni rad i (4) fazi ekspertskog sistema koji služi da određivanje parametara personalizovane terapije. Deo sistema, označen kao LOGOMON (Pentiuc i sar., 2010), namenjen za početnu procenu dece sa govornim poremećajima, njihovu registraciju u bazi podataka, predlog dijagnoze, sa mogućnošću za stručnjaka da potvrdi ili izmeni ovu dijagnozu, upravljanje terapijskim procesom i nadzor napretka dece. Glavni cilj ovog dela sistema je da prikupi snimke dečijeg izgovora, koristeći različita audio okruženja tokom snimanja, u cilju korišćenja nekih fonema za obuku sistema za prepoznavanje u realnom vremenu. Automatska detekcija dislalije, vrši se na nivou foneme i opisana je u referenci Schipor i sar., 2012. Za ispitivanje korišćeni su glasovi /r/,/s/ i /š/ koji su se na osnovu statističkih istraživanja pokazali kao foneme sa najfrekventnijom nepravilnom atrikulacijom. Baza se satoji od 3551 izgovora od kojih su 1428 bila korektna izgovora a 2123 nepravilna. Ocene su davane od strane eksperata i bile su u opsegu 0 do 3 gde je 0 predstavljala neprepoznatljiv, 1 prepoznatljiv ali loš, 2 srednji i 3 zadovoljavajući izgovor. Stimulusi su parametrizovani MFCC koeficijantima, dok je HMM je korišćen za modelovanje izgovora. Odlučivanje se baziralo na verovatnoći da je data observacija generisana određenim modelom. Rezultati su pokazali 54 % poklapanja automatske i

ekspertske ocene (intra ekspertska korelacija iznosila je 0.84 a interekspertska 0.76). Dodatno, razvijen je i pomoćni softver za segmentaciju na nivou fonema (Schipor i Nestor, 2007).

Jedan pristup terapiji okluzivnog sigmatizma, dat je u Benselama i sar. (2007). Ispitivanja su vršena na fonemama /s/ i /š/ gde je nepravilan izgovor nastajao tako što je jezik iz ispravnog, post-alveoralnog položaja, pomeren u alveolarni, a ponekad i dentalni položaj rezultirajući nepravilnom artikulacijom. MFCC koeficijenti su korišćeni za parametrizaciju govornog signala, HMM i ANN su korišćene za autmatsku segmentaciju i procenu devijacije izgovora na nivou foneme. U iterativnom postupku, pacijentima je prezentovan audio i video materijal kako sa pravilnim izgovorima i položajima artikulacionih organa tako i sa sopstvenim, nepravilnim, izgovorom. Tokom terapije pokazano je da ovaj pristup daje dobre rezultate i da se patološki izgovor posmatranih fonema približava normalnom izgovoru.

U Valentini-Botinhao (2012) predstavljen je sistem za automatsku detekciju sigmatizma. Baza stimulusa satojala se od dve grupe izgovora. Prvu grupu stimulusa činili su normalan i patološki izgovor 39 odraslih osoba kod kojih nije dijagnostifikovana patologija govora. Oni su patološke govorne stimuluse izgovarali tako što su simulirali jedan od tri oblika sigmatizma (interdentalni, koronalni i lateralni). Drugu grupu stimulusa činili su snimci izgovora šestoro dece uzrasta od 7 do 18 godina, kod kojih se sigmatizam javio kod tri ispitanika, i jedne odrasle osobe, takođe sa sigmatizmom. Posmatrani su izolovani fonemi /s/ i /z/ u različitim položajima u okviru reči. Na osnovu analize patoloških izgovora u spektralnom domenu, kao obeležja govornih uzoraka za dalju obradu izdvojene su četiri grupe parametra: energija signala u oblasti od 5-11 kHz i 11-20k Hz, 24 MFCC koeficijenta (12 statičkih i 12 dinamičkih), Supervektor koji se sastojao od više parametara modela Gausovih smeša čiji su parametri estimirani *Maksimum A Posteriori* metodom i uprošćeni Supervektor. Korišćena su četiri različita klasifikatora iz *WEKA toolbox* - a (Witten i Frank, 2005) i to *NaiveBayes*, *ViaRegression*, *SVM* i *AdaBoostM1*. Za prvu grupu stimulusa (bazu normalnih i simuliranih patoloških izgovora), koristeći metodu *Leave-one-Out* za obuku i testiranje, najbolji rezultati prepoznavanja patologije iznosili su 86 %, 87% i 94% na nivou foneme, reči i govornika, redom. Kod druge grupe ispitanika, poznavanje patologije postignuto je sa tačnošću od 75,7%, 68,75% i 62,5% na nivou foneme, reči i

govornika. Što se tiče klasifikacije tipa sigmatizma postignut je rezultat od 63% na nivou govornika.

Griogore i sar. (2010) i Velican i sar. (2012) posmatrali su rotacizam, i predložili algoritamsku proceduru kojom bi se mogao detektovati korektan i nekorektan izgovor konsonanta /r/. Govorna baza se sastojala od izgovora reči kod kojih je ispitivani glas bio u inicijalnoj poziciji. Govornici su bili 3 muške osobe (jedna sa rotacizmom) i 8 ženskih (3 sa rotacizmom). Za parametrizaciju su korišćeni MFCC koeficijenti, a za klasifikaciju kNN. Eksperimenti su pokazali da se na zadatom skupu stimulusa optimalni rezultati dobijaju za 11 suseda za kNN sa tačnošću od 78%. Povećavajući bazu ispitanika i koristeći *Walsh-Hadamard Transform* (WHT) rezultati su pokazali tačnost i do 92.5%.

4 Sistem za detekciju patologije govora

Tradicionalne metode koje se primenjuju u terapiji govora i jezika baziraju se na direktnoj interakciji između pacijenta i terapeuta i podrazumevaju niz aktivnosti koje su razvijene u okviru logopedske nauke i prakse, a sa ciljem postavljanja dijagnoze i sprovođenja tretmana. Direktna interakcija je neophodna i efikasna (nezamenjiva) sa aspekta formiranja stručnog mišljenja terapeuta i uspostavljanja odnosa terapeut pacijent. Međutim, ovakav pristup zahteva pravovremeno angažovanje velikog broja terapeuta kako bi se pomoglo svim potencijalnim pacijentima, prevashodno u školama i stručnim institucijama, i time obezbedile mogućnosti za maksimalni napredak i oporavak. Na žalost, potreban broj stručnjaka nije moguće obezbediti u svim slučajevima. Pored toga, ovakva evaluacija je subjektivna i u mnogome zavisi od iskustva terapeuta, njegove trenutne koncentracije, psiho-fizičkog stanja i ostalih ljudskih faktora koji variraju vremenom. Razvojem sistema koji bi automatski vršio procenu poremećaja govora i koji bi se koristio kao ispomoć u dijagnostici i terapiji, ova dva bitna faktora bila bi minimizirana. Na taj način bi se obezbedili uslovi u kojima bi na promene u evaluaciji kvaliteta govora isključivo uticali faktori koji potiču od samog ispitanika. Isto tako, bila bi omogućena testiranja više ispitanika istovremeno, bez potrebe za dolaskom u stručne institucije, što bi umnogome približilo i omasovilo preventivna testiranja patologije govora.

Neki od segmenata dijagnoze i terapije mogu biti zamenjeni i podpomognuti postojanjem automatskog sistema za detekciju poremećaja govora. Skrining testovi na prvom mestu doneli bi mogućnost masovnog testiranja bez nadzora stručnog lica i time blagovremeno ukazali na neophodnost detaljnije provere ili odagnali sve sumnje u postojanje poremećaja govora. Time bi se znatno poboljšale šanse za habilitaciju i rehabiliraciju kod osoba sa poremećajima govora, uz značajnu uštedu materijalnih i ljudskih resursa. Neki od repetitivnih postupaka koji se javljaju tokom terapije, kao što

su određene reevaluacije koje se praktikuju nakon određenog perioda, bile bi pojednostavljenje, smanjile bi angažovanje terapeuta i omogućile češću primenu ukoliko bi bile izvedene automatski. Pored toga, ovakav automatski i objektivan pristup pomogao bi kod ujednačavanja kriterijuma samih terapeuta kako između terapeuta tako i po pitanju pacijenata. Takođe, prilikom obučavanja terapeuta, sistem kojim se automatski detektuje patologija govora bio bi koristan za samoocenjivanje i kontrolu samog terapeuta. Imajući u vidu navedene razloge i prednosti automatske detekcije, javlja se potreba za postojanjem jednog takvog funkcionalnog sistema za automatsku detekciju patologije u izgovoru.

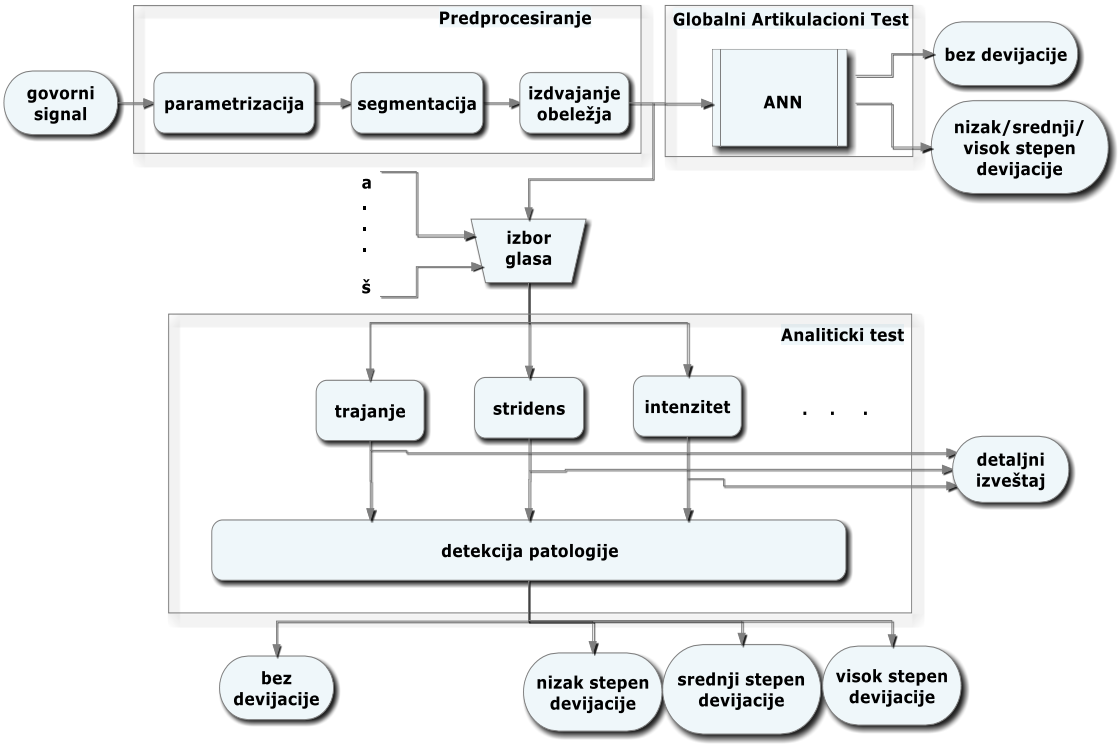
Sistem za automatsku detekciju patologije govora, sa obzirom da se koristi kao pomoć u terapiji i dijagnostici mora se bazirati na procedurama koje su usvojene u logopedskoj praksi. Kao što je objašnjeno u prethodnom poglavlju, postoji dobro razvijen skup testova (Kostić i sar. 1983; Vladisavljavić, 1981) koji se koriste u svrhu procene nivoa govorno-jezičkog razvoja i sposobnosti. Automatski sistem treba da prati navedene procedure i da daje rezultate koji su kompatibilni sa onima dobijenim tradicionalnim metodama.

Treba imati u vidu da se, prilikom formiranja sistema za automatsku procenu patološkog govora i njegovih modula, pomoću algoritamskog (proceduralnog) rešenja teži postizanju istih rezultata koji se dobijaju perceptualnom procenom. U tom smislu, najpre je potrebno izvršiti analizu i sistematizaciju onih karakteristika govornog signala koja doprinose percepciji patologije u govoru i odrediti njihov doprinos i međusobnu vezu. Zbog toga je potrebno izvesti istraživanja svih vrsta patologija koje se mogu javiti i utvrditi njihove manifestacije u akustičkom (vremenskom i frekvencijskom) domenu i na osnovu njih usvojiti odgovarajuće algoritamske procedure i metode digitalne obrade govornog signala.

4.1 Model sistema za prepoznavanje patološkog govora

Jedan od ciljeva ovog rada jeste razvoj sistema za automatsku detekciju artikulacionih poremećaja kod izgovora glasova. Na slici 4.1 prikazana je blok šema predloženog sistema. On se sastoji od tri osnovna modula: modula za predobradu

govornih stimulusa, modula za realizaciju globalnog artikulacionog testa i modula za realizaciju analitičkog testa. Svakom od modula odgovara specifično softversko rešenje u okviru koga su sprovedena opsežna testiranja i istraživanja u cilju postizanja optimalnih karakteristika. U nastavku će biti dat opis sistema dok će u narednim poglavljima biti detaljno opisana sprovedena istraživanja. Sistem je projektovan kao softverski paket koji na ulazu ima snimljeni govorni stimulus, a na izlazu daje kvantitativnu i kvalitativnu ocenu patologije govora.



Slika 4.1 Blok šema predloženog ekspertskog sistema

4.2 Predobrada govornog signala

Predobrada govornog signala je važan prvi korak. Imajući u vidu da ona prestavlja prvi korak u sistemu potrebno je da bude sprovedena pouzdano i sa što većom tačnošću jer greške nastale u ovom koraku propagiraju kroz sistem i mogu značajno

narušiti njegove performanse. Na ulaz ovog modula dovode se reči izgovorene od strane ispitanika, za koje je potrebno odrediti kvalitativnu i kvantitativnu ocenu patologije. Kako ovaj sistem prati tok logopedske terapije i u osnovi ima testove za ocenu artikulacionih poremećaja bazirane na fonemama, nameće se potreba za segmentacijom reči na foneme i ili subfonemske segmente. U okviru predobrade segmentacija predstavlja najbitniji korak. Uslovi pod kojima se segmentacija vrši za potrebe detekcije patologije govora su specifični u tom smislu da se od ispitanika očekuje kooperativnost. Metodologija ispitivanja je takva da se od ispitanika zahteva da izgovara unapred pripremljen skup stimulusa koji mogu biti reči, rečenice i sl. Karakteristika ovog sistema jeste da je reč koja se segmentira unapred poznata, tako da postoji mogućnost adekvatnog modelovanja analizirane reči u toku obrade. Tačnost algoritama za segmentaciju značajno se povećava u slučajevima kad postoji *apriori* znanje o reči koja se segmentira. Za potrebe ovog sistema ispitano je korišćenje različitih načina parametrizacije signala (MFCC koeficijenata, GFCC koeficijenata, frekvencijski spektar), kao i različitih metoda pronalaženja granica segmenata (ML algoritam, dinamičko vremensko usklađivanje, k najbližih suseda) (Bilibajkić i sar., 2007, 2010, 2011).

Koristeći segmente dobijene automatskom segmentacijom, vrši se izdvajanje obeležja signala pogodnih za dalju analizu. Kako je sistem osmišljen tako da prati procedure korišćene u logopedskoj praksi, za ocenu patologije izgovora koriste se dva testa i to (1) Globalni artikulacioni test i (2) Analitički test, opisani u poglavlju 2.

4.3 Modul globalni artikulacioni test

Sistemski modul koji odgovara Globalnom atrikulacionom testu na ulazu ima izdvojeni fonem, u posmatranoj reči, koji se analizira. Na izlazu modula dobija se kvantitativna ocena odstupanja glasa, odnosno, procena stepena odstupanja njegove artikulacije, ukoliko je prisutna. Ovaj softverski modul izveden je pomoću veštačkih neuralnih mreža (ANN - *Artificial Neural Networks*) (Bilibajkić i sar. 2014, Furundžić i sar.,2006, 2007, 2012). Neuralne mreže se odlikuju mogućnošću formiranja znanja na osnovu klasa uzoraka i prepoznavanja ulaznog uzorka na osnovu sličnosti sa obučavajućim uzorcima. Pri tome one pokazuju dobra diskriminativna svojstva između

klasa. Ova karakteristika neuralnih mreža se posebno ističe kada je u pitanju detekcija obeležja koja utiču na kvalitet izgovornog glasa jer je potrebno razdvojiti klase normalnog i patološkog izgovora posmatrajući realizacije istog fonema. Rezultati istraživanja na temu upotrebe veštačkih neuralnih mreža u detekciji globalne ocene artikulacije data je u poglavlju 6.

4.4 Modul artikulacioni test

Globalnim artikulacionim testom ulazni govorni signal pridružuje se odgovarajućoj klasi patološkog, odnosno, normalnog izgovora. Ovakva klasifikacija nije dovoljno detaljna da bi se mogli identifikovati pojedinačni poremećaji artikulacije i ukazali na to šta treba ispraviti u artikulaciji ispitanika. U tom cilju koristi se artikulacioni test (AT). U modulu koji predstavlja artikulacioni test, vrši se algoritamska procena prisustva određenih devijacija u artikulaciji, a prema grupi poremećaja koja odgovara analiziranoj fonemi. Modul analitičkog testa sastoji se od više podmodula, od kojih svaki odgovara određenoj vrsti devijacije. U logopedskoj praksi razlikuje se 37 različitih vrsta devijacija koje se javljaju kod fonema srpskog jezika u zavisnosti od fonemske grupe. Prethodne studije (Punišić i sar., 2007, 2011a; Punišić, 2012) ukazale su na foneme kod kojih se javlja najveća učestalost atipične artikulacije. Rezultati tih studija ukazali su i na to koja su odstupanja najfrekventnija. Pored toga, za neka od odstupanja, utvrđeno je i postojanje akustičkih karakteristika pomoću kojih je moguće izvršiti distinkciju između normalnog i patološkog izgovora. Upravo iz tih razloga, dalja istraživanja na temu automatske detekcije usmerena su u pravcu prepoznavanja patologija za koje postoje teorijske osnove u vidu definisanih parametara za razdvajanje normalnog i patološkog glasa, kao i onih poremećaja artikulacije koja su najzastupljenija. Detaljan opis istraživanja i programskih modula za automatsku detekciju pojedinačnih devijacija dati su u poglavlju 7.

4.5 Baza ispitanika

Kako bi rezultati primenjenih algoritama bili uspešno analizirani, bilo je potrebno oformiti i obraditi bazu ispitanika, odnosno njihovih izgovora. Prilikom formiranja baze uzoraka, posebna pažnja je usmerena na to da se od ispitanika očekuje da imaju automatizovanu artikulaciju svih glasova srpskog, kao maternjeg jezika. Takođe, korišćena baza izgovora formirana je tako da sadrži dovoljan skup uzoraka za obučavanje i testiranje modela percepcije tipičnog i atipičnog izgovora. Pored toga, bazom je obuhvaćen reprezentativni uzorak u pogledu artikulaciono-akustičkih obeležja svih glasovnih grupa koja najčešće karakterišu atipičan izgovor.

Baza je sačinjena na osnovu izgovora odraslih osoba i dece. Usvojeni uzorci izgovora koji pripadaju odraslim osobama formirani su prema izgovoru 48 ispitanika oba pola, starosti od 21 do 42 godine. Deo baze koji se tiče dečijeg govora formiran je na osnovu izgovora 410 dečaka i devojčica uzrasta između deset i jedanaest godina. Svaki od ispitanika je izgovarao zadate stimuluse u formi izolovanih stimulus - reči iz liste stimulusa Globalnog artukulacionog testa, njih 30, što je rezultiralo skupom od 12300 reči - uzoraka dečijeg govora i 1440 uzoraka govora odraslih ispitanika. Ispitanici su ponavljali reči neposredno za ispitivačem, svojim tempom i intenzitetom, u razmaku od 2 do 4 sekunde. Izgovori svih ispitanika su snimani u zvučno izolovanom prostoru (tiha soba), mikrofonom, sa frekvencijom odmeravanja 44,1 kHz i 16 bita AD konverzijom. Snimci su sačuvani u *wav* formatu. Ova govorna baza je segmentirana od strane obučenih eksperata audo-vizuelnom metodom korišćenjem softverskog paketa *Praat* (Boersma, Weenink, 2010). Svi uzorci iz govorne baze su preslušani, sprovedena je analiza po pitanju ocene patologije, a rezultati su statistički obrađeni da bi se utvrdilo prisustvo svih elemenata patologije i njihova zastupljenost. Odluka o kvalitetu atrikulacije svakog pojedinačnog glasa ispitanika donošena je na osnovu postojećeg standarda pravilne artikulacije, odnosno, opisa akustičko-artikulacionih karakteristika izgovornih glasova srpskog jezika datog u referencama Miletić, 1952; Stevanović, 1981; Kostić i sar., 1964. Imajući u vidu da na percepciju utiču i auditorne i vizuelne informacije (Massaro, 1999.), odgovori ispitanika su ocenjivani samo na osnovu auditorne informacije.

5 Segmentacija

Segmentacija govora je bitan korak u mnogim realizacijama sistema za digitalnu obradu govora, kao deo sistema za prepoznavanje jezika, prepoznavanje govornika, sintezu govora, detekciju patologije govora i drugo (Vidal i Marzal, 1990; Ljolje i sar., 1996; Rabiner i Jung, 1994). Takođe, ona se koristi u fazi obuke velikog broja sistema za obradu govornog signala kojima je neophodna relativno velika baza govornih segmenata dobijenih iz kontinualnog govora (Kvale, 1993).

Istraživanja su pokazala da perceptivna analiza kontinualno izgovaranih vokala sama po sebi nije dovoljna za procenu kvaliteta patološkog izgovora (Hammaberg i sar., 1980) jer ne sadrži dinamičke aspekte kontinualnog govora. Kontinualno izgovarani vokali nisu reprezentativni što se konteksta komunikacije tiče jer ne sadrže segmentne i suprasegmentne karakteristike koje utuču na perceptivnu procenu kvaliteta govora (de Krom i sar., 1995). U sistemima predviđenim za automatsku procenu poremećaja govora, a posebno u slučajevima gde se procena vrši na nivou glasova realizovanih u okviru reči i rečenica, bitan korak predstavlja segmentacija reči (Paulraj i sar., 2010; Godino-Llorente i sar., 2009; Bilibajkić i sar., 2010; Schipor i Nestor, 2007). Automatska segmentacija je važan korak jer je jedna od prvih faza obrade i kao takva mora biti adekvatno izvedena jer se greške nastale u ovom koraku propagiraju kroz čitav sistem i time degradiraju njegov kvalitet. Segmentacija koja se koristi u svrhu detekcije patologije govora se može vršiti na nivou reči, odnosno govornog stimulusa koji se ispituje sa ciljem da se izdvoji samo govor pacijenta (Schipor i Nestor., 2007; Maier i sar., 2009), kao i na nivou fonemskih i subfonemskih segmentata ispitivanog govora (Bilibajkić, 2011; Paulraj i sar., 2010; Fredouille i sar.,2011; Godino-Llorente i sar., 2009). Metode koje se primenjuju za automatsku segmentaciju govora mogu biti klasifikovane prema raznim kriterijumima (Vidal i Marzal, 1990). Jedan od osnovnih kriterijuma je klasifikacija prema tome da li postoji lingvistički opis korpusa koji se

segmentira ili ne. Prema tom kriterijumu razlikujemo slepu (engleski *blind*) i poluautomatsku (engleski *aided*) segmentaciju. U slučajevima kada je metodologija ispitivanja takva da se od ispitanika traži da izgovara unapred pripremljen govorni material u vidu slogova, reči, rečenica i sl. rešava se problem poluautomatske segmentacije i postoji mogućnost adekvatnog modelovanja analiziranog stimulusa u toku obrade. Tačnost algoritama za segmentaciju značajno se povećava kad postoji *apriori* znanje o stimulusima koji se segmentiraju.

U nastavku su data dva primera segmentacije reči optimizovana za patološki govor. Kao što je prikazano na slici 5.1, bazira se na stimulusima definisanim *Globalnim artikulacionim testom.* Sa tim u vezi od ispitanika se očekuje kooperativnost u smislu da reči izgovara redom kojim su u testu navedene. Ono što je specifično za takav algoritma jeste to da on mora biti robustan na sve varijetete izgovora koji potiču od mogućih patologija, godina starosti, emocionalnog stanja, polne pripadnosti ispitanika i sl. U poglavlju 5.1 je predstavljen postupak za segmentaciju ML algoritmom, a u poglavlju 5.2 je dat prikaz postupaka baziranih na vremenskom usklađivanju.

5.1 Segmentacija bazirana na ML algoritmu

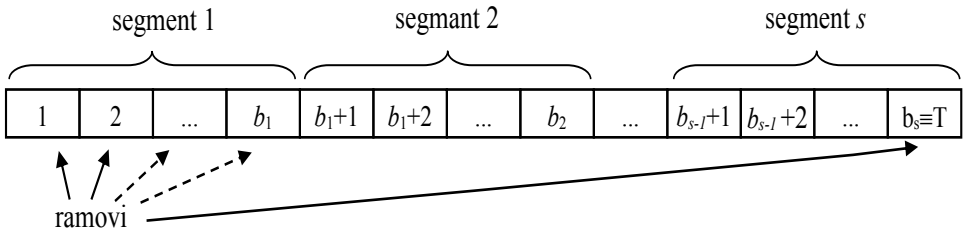
Algoritmi maksimalne verodostojnosti primenjeni na segmentaciju govora (*Maximum Likelihood*) ML (Šarić i sar., 1995; Turajlić i sar., 1993; Vidal i Marzal, 1990; Swedsen i Soong, 1987) pokazuju dobre rezultate kada se koriste za izdvajanje stacionarnih segmenata govornog signala, međutim, njihova preciznost opada u slučajevima naglih tranzicija niskog intenziteta. Sa druge strane, kada se koriste sekvencijalni algoritmi (Vidal i Marzal, 1990; Šarić, 1994; Andre-Obrecht i sar., 1988; Appel i sar., 1983)postiže se visoka tačnost detekcije naglih promena u spektru i energiji signala, dok se u slučajevima kada je promene spektra kontinualna njihova preciznost smanjuje.

Imajući u vidu navedene karakteristike dva pristupa segmentaciji govora, ML algoritam se može unaprediti tako da sa visokom preciznošću određuje položaj naglih tranzicija (kao što to mogu sekvencijalni algoritami), zadržavajući pri tome visoku

pouzdanost razdvajanja stacionarnih segmenata sa sporom promenom spektra i snage signala. Predloženi algoritam se bazira na generalizovanoj verodostojnosti i Itakurinoj meri odstojanja za AR (auto regresivni) model.

5.1.1 ML algoritam

Digitalizovani govorni signal x_t deli se na T nepreklapajućih ramova jednakih dužina od N odbiraka. Dužina ramova se bira tako da ramovi budu dovoljno kratki da se može smatrati da su generisani jedinstvenim AR modelom, a dovoljno dugački da se na njima može proceniti AR model. Cilj je da se ramovi grupišu u segmenate tako da promena energije i spektra unutar svakog od segmenata bude minimalna. Kako metodologija ispitivanja dozvoljava da se unapred postavi broj segmenata, smatraćemo da je on poznat i da je jednak s. Segmantaciju sada možemo posmatrati kao pronalaženje s ramova koji predstavljaju krajeve, odnosno granice segmenata tako da je ispunjen zadati kriterijum (minimalne promene energije i spektra). Kao što je prikazano na slici 5.1, sa b_i je označen indeks poslednjeg rama segmenta i, pa se segmentacija svodi na nalaženje niza indeksa $\{b_1, b_2,..., b_s\}$.



Slika 5.1 ML segmentacija

Sa B_s ćemo označiti jedan od mogućih položaja granica segmenata $B_s = \{b_1, b_2,..., b_s\}$.

Ako sa $L(s_i \mid \hat{\theta}_i)$ označimo logaritam verodostojnosti da je segment s_i generisan jedinstvenim AR modelom sa parametrima $\hat{\theta}_i$ estimiranim na segmentu s_i, tada se $L(s_i \mid \hat{\theta}_i)$ prema (Itakura, 1975; Appel i sar., 1983), izražava relacijom

$$L(s_i \mid \hat{\theta}_i) = -l_i \log\left(\sigma^2(b_{i-1}+1, b_i)\right) \tag{5.1}$$

gde je l_i ukupan broj segmenata s_i, a $\sigma^2(b_{i-1}+1, b_i)$ je srednja kvadratna greška linearne predikcije za AR model estimiran na segmentu s_i. Ispunjenost kriterijuma maksimalne stacionarnosti segmenata ekvivalentan je maksimiziranju logaritma verodostojnosti segmentacije $L(B_s)$ što se izražava relacijom (Vidal i Marzal, 1990; Swedsen i Soong, 1987)

$$L(B_s) = \sum_{i=1}^{s} L(s_i \mid \hat{\theta}_i) \tag{5.2}$$

Unutrašnja distorzija segmenta s_i, označena je sa $d(b_{i-1}+1, b_i)$ i predstavlja meru odstojanja ramova unutar segmenata. U slučaju primene AR modela, za meru distorzije segmenta obično se usvaja negativna vrednost logaritma verodostojnosti $L(s_i \mid \hat{\theta}_i)$ kao u relaciji (5.3)

$$d(b_{i-1}+1, b_i) = -L(s_i \mid \hat{\theta}_i) = -l_i \log\left(\sigma_i^2(b_{i-1}+1, b_i)\right) \tag{5.3}$$

Ukupna distorzija segmentacije $D(B_s)$, u slučaju AR modela ima oblik

$$D(B_s) = -L(B_s) = \sum_{i=1}^{s} d(b_{i-1}+1, b_i) \tag{5.4}$$

Raspored granica B_s, sa najmanjom ukupnom distorzijom $D(B_s)$ predstavlja segmentaciju sa najvećom verodostojnošću. Da bismo odredili traženi raspored granica, potrebno je od svih mogućih podela T ramova na s segmenata izabrati onu B_s sa minimalnom ukupnom distorzijom $D(B_s)$ (Šarić i sar., 1995; Turajlić i sar., 1993; Swedsen i Soong, 1987). Ako se za nalaženje minimuma kriterijumske funkije $D(B_s)$ koristi relacija:

$$D(\overline{B}_{i+1}) = \min_{b_i} \left\{ D(\overline{\{b_1, \ldots, b_i\}}) + d(b_i + 1, b_{i+1}) \right\} \tag{5.5}$$

gde je sa $D(\overline{B}_i)$ označena ukupna distirzija optimalne segmentacije $\overline{B}_i$, tada je moguće, na osnovu osobina izraženih relacijom (5.5) primeniti algoritam višenivoskog dinamičkog programiranja prilikom nalaženja minimuma ukupne distirzije $D(\overline{B}_s)$ (Swedsen i Soong, 1987).

5.1.2 Isticanje naglih promena

Eksperimentalno je utvrđeno da ML agoritam favorizuje promene velikog intenziteta što uzrokuje greške u određivanju brzih promena malog intenziteta. Da bi rezultati segmentacije bili verniji fonetskom sadržaju reči, potrebno je povećati osetljivost algoritma na nagle pomene niskog intenziteta. Iz tog razloga je kriterijumskoj funkciji datoj u (5.4) dodat još jedan član kojim se određuje verovatnoća prisustva nagle promene spektra na granici dva segmenta. Relacijom (5.6) dat je logaritam odnosa verodostojnosti promene AR modela na pretpostavljenoj granici segmenta s_i

$$L_c(b_i) = 2N\log(\sigma^2(b_i : b_i + 1)) - N\log(\sigma^2(b_i)) - N\log(\sigma^2(b_i + 1)), \quad (5.6)$$

gde je sa $\sigma^2(j:k)$ označena srednje kvadratna greška linearne predikcije na ramovima od indeksa j do k. Da bi se dobila „mekana“ odluka o prisustvu nagle promene spektra, definisana je funkcija $\Psi(b_i)$ kao:

$$\Psi(b_i) = \log\left(\max\{\psi(b_i), 1\}\right), \quad (5.7)$$

gde je niz $\psi(k)$ mera lokalne konveksnosti vremenske serije $L_c(k)$ dobijena relacijom:

$$\psi(k) = L_c(k) - \left(L_c(k-1) + L_c(k+1)\right)/2 \quad (5.8)$$

Finalna kriterijumska funkcija za segmentaciju govornog signala ima oblik:

$$\tilde{D}(B_s) = \sum_{i=1}^{s}\left(d(b_{i-1}+1, b_i) - \alpha\Psi(b_{i-1})\right), \quad (5.9)$$

gde je konstanta α faktor kojim se podešava uticaj nagle promene u spektru signala u odnosu na kriterijum minimalne distorzije segmenta. Kao i u slučaju kriterijumske funkcije $D(B_s)$, funkciju $\tilde{D}(B_s)$ minimizirano algoritmom višenivoskog dinamičkog programiranja smatrajući faktor $-\alpha\Psi(b_{i-1})$ lokalnom težinom čvora.

5.1.3 Eksperimentalni rezultati

Kvantitativna ocena kvaliteta segmentacije izvršena je poređenjem rezultata segmentacije dobijenih pomoću dva algoritma. U prvom slučaju korišćen je ML algoritam dok je u drugom slučaju korišćen ML algoritam sa pojačanom osetljivošću na

nagle promene spektra. Testiranje je izvršeno na stimulusima govorne baze Globalnog Artikulacionog Testa. Učestanost odabiranja govornog signala decimacijom je sa 44.1 kHz svedena na 11.025 kHz. Korišćen je auto regresivni model reda 10 a dužina pojedinih ramova bila je 10 ms. Za tačan položaj granica fonema uzete su one dobijene od strane eksperata. Granice su ekspertski utvrđene na osnovu vremenskog dijagrama, spektralne analize i auditivne kontrole pojedinih fonetskih celina. Analiziran je izgovor sedam ispitanika iz dela baze izgovorenog od strane odraslih osoba na osnovu izolovano izgovorenih reči /mama/, /baba/, /žaba/ i /riba/. U tabeli 5.1. dat je prikaz rezultata segmentacije i u obliku ispravno postavljenih granica, ispuštenih granica fonema i ubačenih nepostojećih granica fonema.

Tabela 5.1 Rezultati segmentacije ML i modifikovanim ML algoritmom

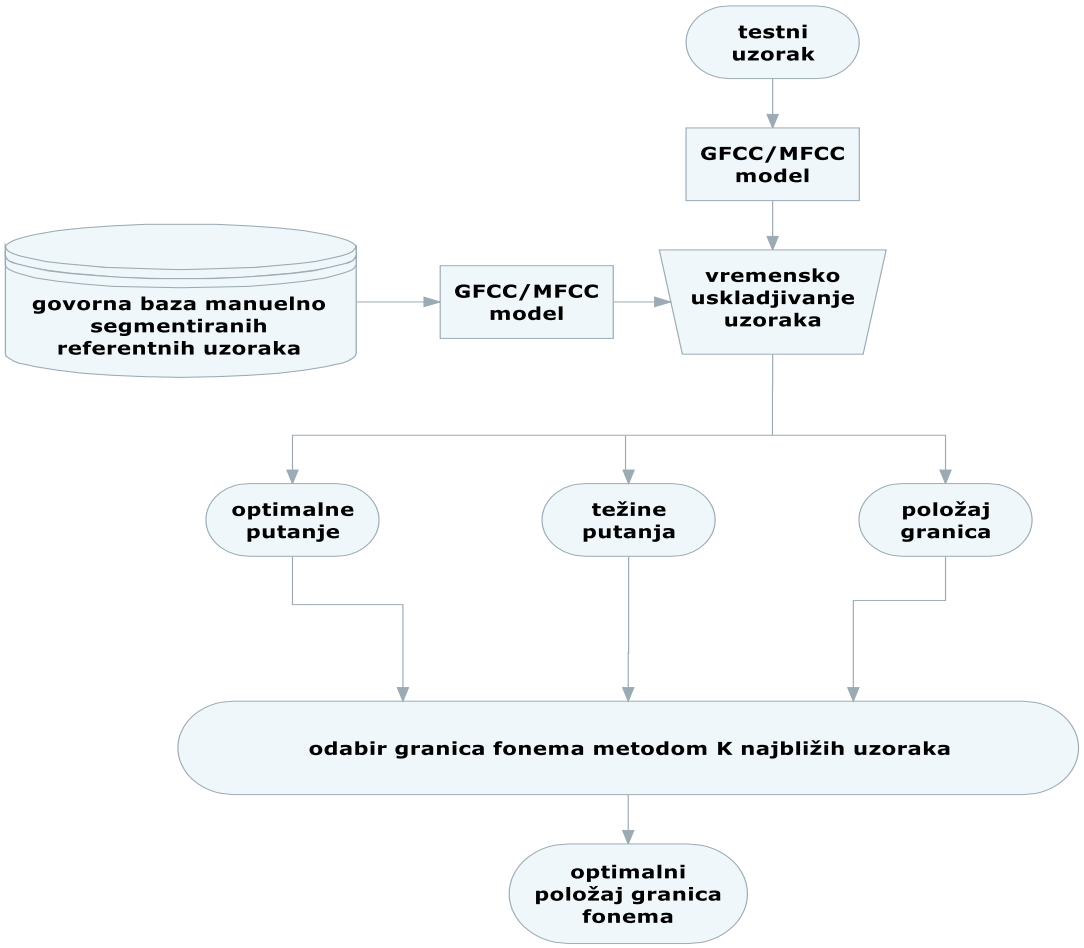
Algoritam	Broj fonema	Ispravno određene granice [%]	Ispuštene granice fonema [%]	Ubačene nepostojeće granice fonema [%]
ML	112	70.2	29.8	39.5
Modifikovani ML	112	75.0	25.0	38.1

5.2 Segmentacija reči pomoću dinamičkog vremenskog usklađivanja

Neki od pristupa koji se najčešće koriste kod prepoznavanja govora, usvojeni su i kao metode segmentacije govornog signala. Naime, za segmentaciju se uspešno koriste skriveni Markovljevi modeli (*Hidden Markov Model* - HMM) (Ljolje i sar., 1996; Angelini i sar., 1997; Farhat i sar., 1993; Chou i sar., 1996; Toledano i sar., 2003; Huggins-Daines i sar., 2006), veštačke neuralne mreže (*Artificial Neural Network* - ANN) (Karjalainen i sar., 1998), kao i hibridni modeli (Malfrere i sar., 1996). Takođe, pored navedenih, a specijalno u sistemima prilagođenim konkretnoj nameni, u svrhu segmentacije, efikasno se koristi i dinamičko vremensko usklađivanje (*Dynamic Time Wrapping*- DTW) (Veprekt i sar., 1996; Cosi i sar., 1991; Paulo i sar., 2003). Razloge za ovo možemo naći u činjenici da se ono lako implementira u manjim sistemima koji imaju ograničenu procesorsku moć kao i u tome da je obučavanje ovih sistema jednostavnije i brže u poređenju sa onima koje koriste HMM i ANN.

U ovom radu DTW se koristi kako bi se izvršilo usklađivanje segmentiranog govornog signala sa nesegmentiranim govornim signalom koji imaju istu fonetsku strukturu. DTW se bazira na poređenju stimulusa pomoću odgovarajuće distantne metrike posle čega obično sledi računanje ukupne akumulirane distance između prototipa i mogućih kandidata. Rangiranje se vrši prema sličnosti sa prototipom.

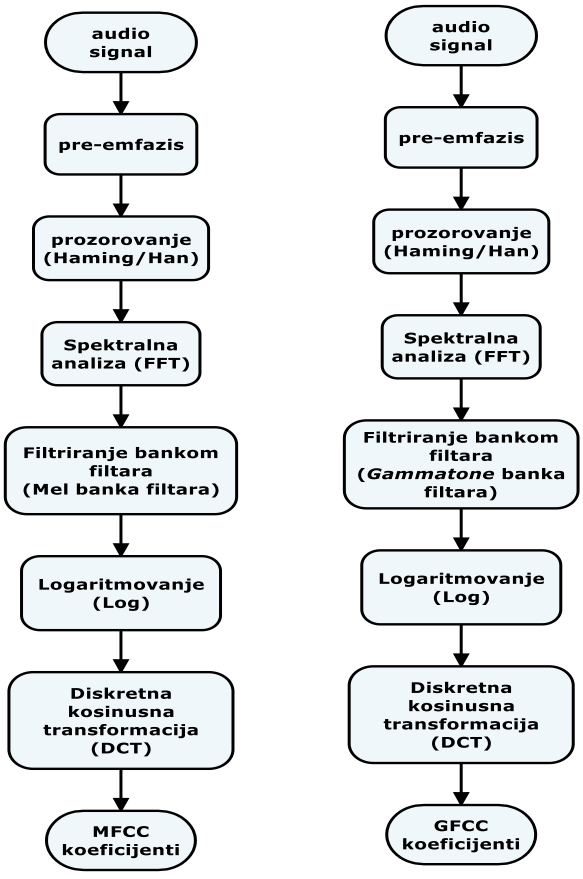
Blok šema postupka za segmentaciju primenom DTW i *k* najbližih suseda (*kNN - k Nearest Neighbors*) prikazana je na slici 5.2. Testni uzorak, odnosno, uzorak kome se žele postaviti granice fonema i subfonema, najpre se parametrizuje kako bi se iz vremenskog domena preveo u domen parametara pogodnih za dalju obradu. Nakon parametrizacije testnog uzorka vrši se njegovo vremensko usklađivanje pomoću DTW sa svim reprezentativnim uzorcima iz baze GAT. Ovim se za testni uzorak dobija niz optimalnih putanja, njihovih težina i položaja granica. Metodom *k* najbližih suseda vrši se izbor optimalnih položaja granica fonema i subfonema.



Slika 5.2 Blok šema predloženog algoritma za segmentaciju

5.2.1 Parametrizacija

U ovom slučaju korišćena su dva vida parametrizacije i to MFCC i GFCC (*Gammatone Frequency Cepstral Coefficients*) koeficijenti. U poslednjih desetak godina Mel frekvencijski kepstralni koeficijenti (MFCC) dominiraju na polju obrade govora. Korišćenje ovih koeficijenata može se smatrati jednim od standarda za izdvajanje obeležja govornog signala. Sa druge strane, inspirisani auditornim modelima percepcije govora za izdvajanje karakteristika govornog signala koriste se parametri koji oponašaju auditorni sistem čoveka. U procesu slušanja, govorni signal prolazi kroz niz promena u auditornom sistemu. Poznato je da se kohlea, koja predstavlja najbitniju komponentu u unutrašnjem uvu, ponaša kao banka filtara i na taj način iz zvučnog signala izdvaja bitne karakteristike. *Gammatone* banka filtara (Holdsworth i sar., 1988), na kojoj su bazirani GFCC oponaša selektivnu funkciju bazilarne membrane. Na slici 5.3. prikazan je proces formiranja MFCC (Davis i Mermelstein, 1980) i GFCC (Shao i sar., 2009) parametara.

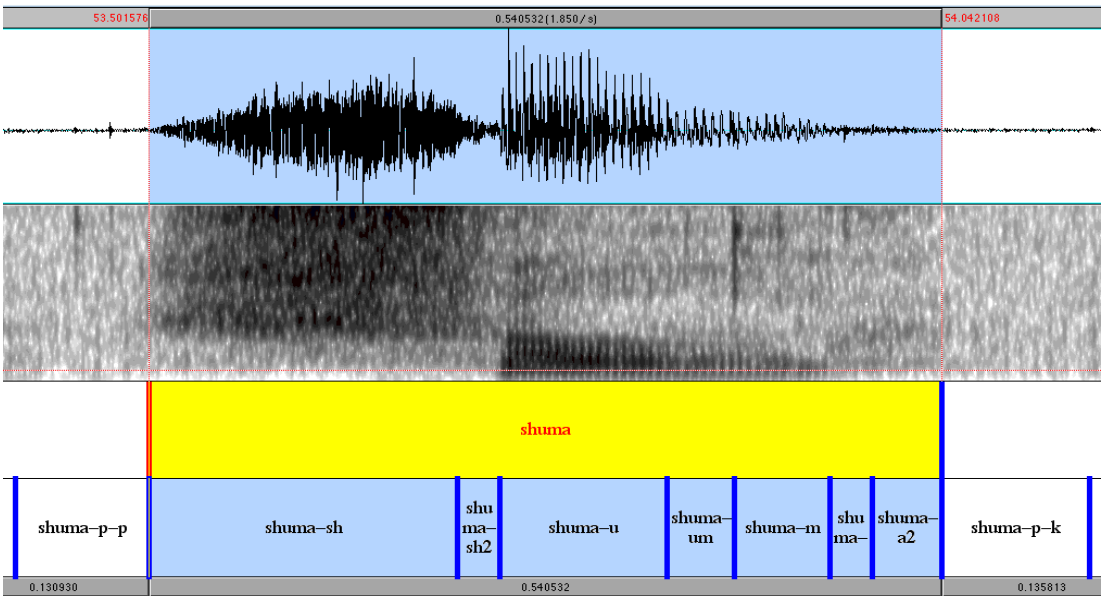


Slika 5.3 Formiranje MFCC i GFCC koeficijenata

Ovi koeficijenti baziraju se na Furieovoj transformaciji i na diskretnoj kosinusnoj transformaciji. Signal se najpre prozoruje Hamingovom ili Haningovom prozorskom funkcijom, nakon čega se računa spektar za svaki vremenski prozor pomoću FFT-a. Zatim se spektar filtrira Mel bankom filtara, odnosno *Gammatone* bankom filtara da bi se formirali Mel frekvencijski koeficijenti, odnosno, *Gammaton*e frekvencijski koeficijenti. Posle toga se vrši logaritmovanje, nakon čega sledi diskretna kosinusna transformacija. Obično se na kraju zadržava samo određeni broj koeficijenata od značaja.

5.2.2 Predobrada govorne baze

Svi varijeteti patološkog izgovora sadržani u bazi obrađeni su na taj način da su granice fonemskih i subfonemskih segmenata unapred određene postupkom ekspertskog slušanja, a na osnovu vremenske predstave i spektrograma govornog signala. Kontrola je vršena auditivnim putem. Manuelno su segmentirane pauze na početku i kraju reči, koartikulacije fonema kao i relaksacije vokala. Prikaz izbora segmenata za reč /šuma/ date su na slici 5.4. Segmentacija je vršena koristeći programski paket PRAAT, koji je pogodan za ovu vrstu obrade jer postoji mogućnost postavljanja granica kao i pregled frekvencijskog sadržaja signala.

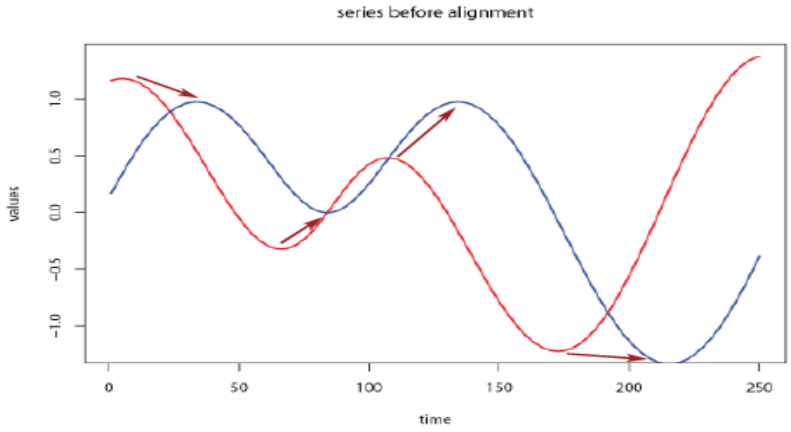


Slika 5.4 Segmentacija reči /šuma/.

5.2.3 Segmentacija uz pomoć DTW algoritma

Dinamičko vremensko usklađivanje je korišćeno kako bi se nesegmentirani testni signal poredio sa svakim od modela u bazi. Na taj način, za svaki par koji čine testni uzorak i uzorak iz baze, dobijan je niz vrednosti za optimalnu putanju, težine putanja i položaje granica.

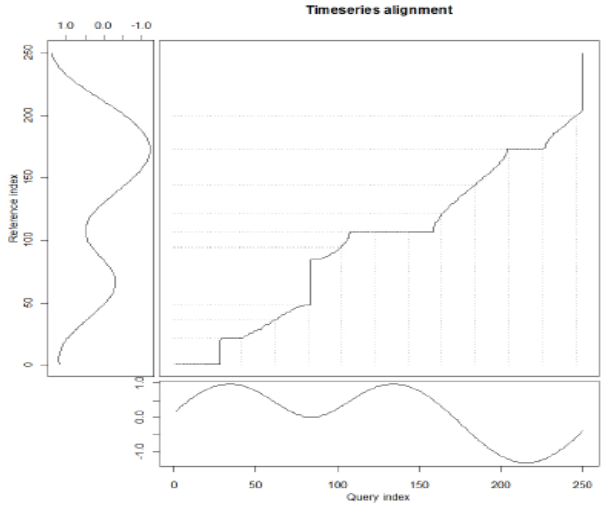
DTW podrazumeva korišćenje dinamičkog programiranja kako bi se vremenski uskladile sekvence na taj način da definisana distanca između njih bude minimalna. Ovaj algoritam pokazao se efikasnim u slučajevima gde je bilo potrebno ustanoviti stepen sličnosti signala a da se pri tom minimizira efekat pomeranja i distorzije po vremenskoj osi (slika 5.5.). Posebno, kod korišćenja u obradi govora, DTW ima uspeha zahvaljujući činjenici da je pomoću njega moguće pronaći sličnosti uprkos vremenskim varijacijama i varijacijama u izgovoru.



Slika 5.5 Vremensko usklađivanje signala

Problem vremenskog usklađivanja možemo definisati na sledeći način. Ako sa $X=(x_1, x_{2,\dots} x_N)$ i $Y=(y_1, y_{2,\dots} y_M)$ predstavimo odbirke u vremenu (slika 5.6.), *DTW* obezbeđuje pronalaženje optimalne putanje vremenskog usklađivanja uzoraka X i Y. Jedino ograničenje jeste da se od signala zahteva da su odbirci uzimani u ekvidistantnim

tačkama na vremenskoj osi. U opštem slučaju, vektori X i Y mogu biti bilo kakve opservacije uzete u ekvidistantnim tačkama.



Slika 5.6 Primer putanje vremenskog usklađivanja

Ako pretpostavimo da observacije signala pripadaju nekom parametarskom prostoru Φ tada, da bismo izvršili poređenje dve sekvence $X, Y \in \Phi$ mora biti definisana lokalna mera odstojanja, data funkcijom

$$d : \Phi \times \Phi \to \Re \geq 0 \tag{5.10}$$

Intuitivno je jasno da *d* ima malu vrednost u slučaju da su sekvence slične, a da je veliko ukoliko su sekvence različite. Funkcija *d* se najčešće naziva funkcijom distance ili distantnom metrikom. U praksi se koriste mnoge mere distance, a dve su date jednačinama:

$$d(i,j) = |x_i - y_i| \tag{5.11}$$

$$d(i,j) = (x_i - y_i)^2 \tag{5.12}$$

Po izboru mere distance, može se formirati matrica distance $C \in \Re^{N \times M}$ koja predstavlja matricu lokalnih distanci

$$C_l \in \Re^{N \times M} : c_{i,j} = \|x_i - y_i\|, \quad i \in [1:N], j \in [1:M] \tag{5.13}$$

Ako sada sa *p* obeležimo niz tačaka koje pripadaju ravni (*i, j*) takve da je

$$p = (p_1, p_2, p_3, \ldots p_K) \tag{5.14}$$

a sa p_l jednu tačku putanje takvu da je

$$p_l = (n_l, m_l) = (i, j)_l \in [1:N] \times [1:M], \qquad l \in [1:K] \tag{5.15}$$

tada p predstavlja putanju vremenskog usklađivanja sekvenci X i Y ukoliko zadovoljava sledeće uslove:

1. granični uslov: p_1 =(1,1) i p_K =(N, M), odnosno, prva i poslednja tačka putanje moraju odgovarati prvoj i poslednjoj tački sekvenci.

2. uslov monotonosti: $n_1 \le n_2 \le ...n_K$ i $m_1 \le m_2 \le ...m_K$. Ovaj uslov obezbeđuje očuvanje vremenske monotonosti.

3. uslov kontinualnosti (veličina koraka): $p_{l+1} - p_l \in \{(1,1),(0,1),(1,0)\}$. Ovim uslovom obezbeđuje se da koraci koji se prave prilikom pronalaženja putanje ne budu veliki.

Težina putanje p tada je data kao:

$$c_p(X,Y) = \sum_{l=1}^{K} c(x_{n_l}, y_{m_l}) \tag{5.16}$$

Ona putanja čija težina ima minimalnu vrednost predstavlja optimalnu putanju vremenskog usklađivanja uzoraka X i Y. Ako tu putanju obeležimo sa P^*, onda važi da je

$$c_{P^*}(X,Y) = \min\{\sum_{l=1}^{K} c_p(x_{n_l}, y_{m_l})\}, p \in P^{N \times M} \tag{5.17}$$

gde je $P^{N \times M}$ skup svih mogućih putanja.

Dakle, problem dinamičkog usklađivanja uzoraka možemo predstaviti kao problem izbora putanje vremenskog usklađivanja uzoraka koja ima minimalnu akumuliranu distancu prema izabaranoj meri odstojanja.

Kod algoritma dinamičkog programiranja problem pronalaženja putanje najmanje težine rešava se formiranjem matrice akumuliranih težina, odnosno matrice globalnih težina D.

Prva kolona matrice D definisana je kao

$$D(1,j) = \sum_{k=1}^{j} c(x_1, y_k), j \in [1, M] \tag{5.18}$$

prva vrsta kao

$$D(i,1) = \sum_{k=1}^{i} c(x_k, y_1), i \in [1, N] \tag{5.19}$$

a ostali elementi matrice su dati rekurzijom

$$D(i,j) = c(x_i, y_j) + \min\{ D(i-1, j-1), D(i-1, j), D(i, j-1)\} \sum_{k=1}^{i} c(x_k, y_1) \qquad (5.20)$$

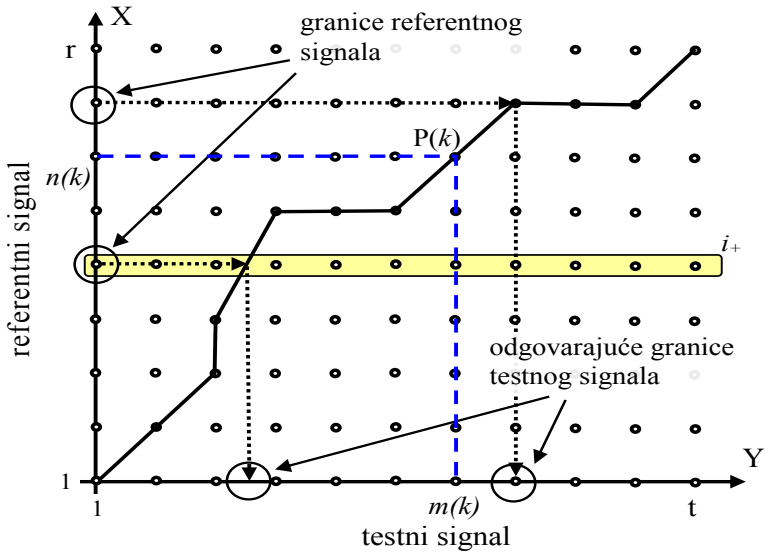
Dakle, u određenoj tački sa koordinatama *(i,j)* vrednost elementa matrice *D(i,j)* računa se kao zbir vrednosti distance odgovarajućih elemenata niza *X* i *Y* i minimuma kumulativnih distanci susednih elemenata. Algoritmom dinamičkog programiranja formira se matrica kumulativnih distanci. Optimalna putanja vremenskog usklađivanja nalazi se na taj način što se od tačke, $p_{end} = (N, M)$ kreće ka tački $p_1=(1,1)$ prateći tačke sa najmanjom kumulativnom distancom u matrici *D*.

Za lokalnu meru odstojanja prilikom poređenja referentnog i testnog signala kod vremenskog usklađivanja uzet je kvadrat euklidskog odstojanja ramova analize dat kao:

$$\mathrm{d(i,\ j)} = \sum_{\mathrm{k=1}}^{\mathrm{p}} \left(G_k^i - G_k^j\right)^2 \qquad (5.21)$$

gde su sa G_k^i i G_k^j dati kepstralni koeficijenti (M*FCC ili GFCC* u zavisnosti od korišćane parametrizacije) referentnog i testnog uzorka, redom. Sa *p* je obeležen broj koeficijenata koji je korišćen u parametrizaciji.

Težina optimalne putanje data je kao suma pojedinačnih distanci duž optimalne putanje dobijene algoritmom vremenskog usklađivanja. Položaj granica fonema testnog signala računat je preslikavanjem granica referentnog signala u odnosu na optimalnzu putanju, kao što je prikazano na slici 5.7, za svaki od modela u bazi.



Slika 5.7 Postavljanje granica

Eksperimentalno je utvrđeno da segmentacija koja se dobija usvajanjem granica dobijenih pomoću referentnog modela sa najmanjom vrednošću težine optimalne putanje nije nužno najbolja (Bilibajkić, 2011). Uzrok tome je postojanje greške procene raspodeljene oko tačne vrednosti. Finalni položaj granica dobija se na osnovu položaja odgovarajućih granica K najbližih suseda. kNN postupak u ovom slučaju nije upotrebljen striktno za klasifikaciju uzoraka. Postupak je modifikovan tako da se pomoću njega određuje K najbližih suseda čiji se doprinosi skaliraju na taj način da susedi koji imaju manju vrednost mere odstojanja od testnog uzorka imaju veći uticaj na dodeljenu osobinu, a oni koji imaju veću vrednost odstojanja, manji uticaj na dodeljivanje osobine.Težina optimalne putanje uzeta je kao kriterijum za izbor suseda. U konkretnom slučaju, utvrđeno je da je za optimalnu vrednost granica potrebno uzeti položaj granica 10 uzoraka sa najmanjim vrednostima distance. Doprinos suseda skaliran je sa koeficijentom C_i, datim kao:

$$C_i = {}^{1}\!\!\diagup\!\!_{S_i}, i = 1...K \tag{5.22}$$

gde je K broj suseda a S_i mera odstojanja i-tog suseda, koja je u ovom konkretnom slučaju predstavljena težinom optimalne putanje, predhodno datom kao $c_{p^*}(X,Y)$. Drugim rečima, finalni položaj granica $q = (q_1, q_2, ...q_k)$ dobija se kao

$$q_k = \frac{\sum_{i=1}^{10}(C_i * p_{i,k})}{\sum_{i=1}^{10} C_i}, \quad k = 1,2...Ng \tag{5.23}$$

gde je k redni broj granice, N_g ukupan broj granica, a $p_i = (p_{i,1}, p_{i,2}, ...p_{i,k})$ su granice testnog uzorka dobijene preslikavanjem granica i-tog suseda pomoću optimalne putanje (slika 5.7).

Predloženi algoritam za segmentaciju izrađen je u programskom paketu MATLAB.

5.2.4 Rezultati istraživanja

U oba slučaja ramovi analize obuhvatali su 256 odbiraka signala, a preklapanje ramova bilo je 50%. Banka filtara sastojala se od 24 filtra. Frekvencijski opseg koji obuhvataju banke filtara je od 110 Hz do 5512 Hz. Konačno, parametrizacija svakog od ramova analize vršena je sa 12 kepstralnih koeficijenata na taj način što je izuziman prvi kepstralni koeficijent, a u obzir su uzimani koeficijenti od prvog do trinaestog. Prilikom rešavanja problema dinamičkog programiranja korišćena su globalna i lokalna ograničenja putanje. Globalna oblast putanje je unapred definisana kao poligon oko glavne dijagonale dok je dozvoljeni nagib putanje zadat u granicama 1/3 do 3. Testiranja su vršena na uzorku izgovora 48 ispitanika. Korišćena su tri stimulusa iz baze GAT i to reči /mama/, /baba/ i /šuma/. U tabeli 5.2 dat je broj uzoraka u zavisnosti od ocene patologije na GAT-u

Tabela 5.2 Raspodela ocena među uzorcima

Ocena na GAT-u	Broj uzoraka
1	0
2	3
3	82
4	48
5	10
6	1
7	0

Testiranje je vršeno *Leave One Out* (LOO) unakrsnom validacijom (Kohavi 1995; Cawley i sar., 2003), odnosno, obučavajući skup činili su svi uzorci baze izuzimajući testni uzorak. Granice dobijene automatskom segmentacijom poređene su sa granicama dobijenim ekspertskom metodom. Ukoliko je greška segmentacije bila manja od 3 rama analize (35 ms) smatrano je da su granice korektno određene. Izuzetak su bile granice na kraju reči kod kojih je tolerancija bila veličine 5 ramova analize. Objašnjenje za ovo treba tražiti u činjenici da su snimci načinjeni u prostoriji koja nije

bila akustički obrađena i u kojoj je, shodno tome, postojala određena reverberacija, tako da je snaga signala na kraju reči opadala sporo te ni od strane eksperata nije bilo moguće tačno odrediti kraj reči.

5.2.4.1 Segmentacija na bazi MFCC modela

Prosečna greška segmentacije je 0.88% dok je greška za pojedinačne slučajeve data u tabeli 5.3.

Tabela 5.3 Greška segmentacije MFCC model

Reč	Šuma	Mama	Baba
Prosečna greška	1.90%	0.42%	0%

U drugom slučaju, kada je za toleranciju greške segmentacije uzeta vrednost od 22.6 ms (dva bloka obrade) prosečna greška segmentacije iznosila 2.26%, a za reči /šuma/,/mama/ i /baba/ data je u tabeli 5.4.

Tabela 5.4 Greška segmentacije MFCC model – strožiji uslov

Reč	Šuma	Mama	Baba
Prosečna greška	4.43%	0.83%	0.83%

5.2.4.2 Segmentacija na bazi GFCC modela

Prosečna greška segmentacije dobijena ovim postupkom bila je 2.22%, dok je za pojedinačne reči data u tabeli 5.5.

Tabela 5.5 Greška segmentacije GFCC model

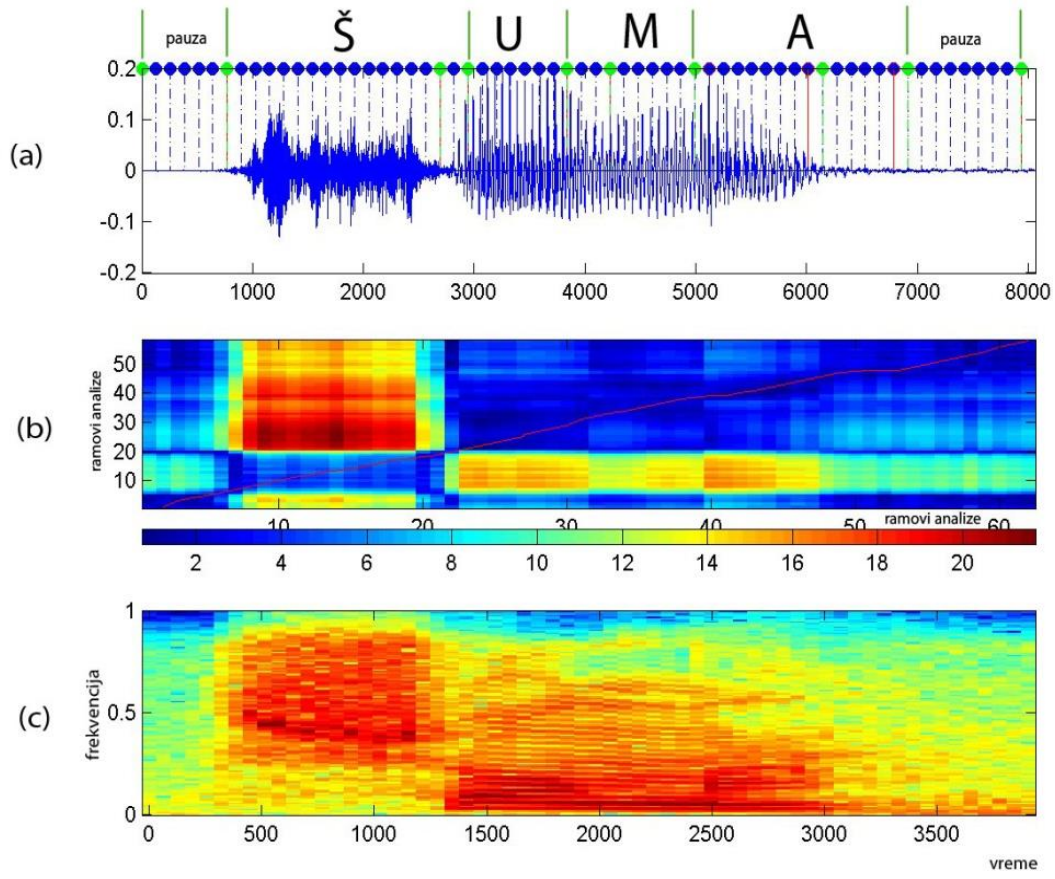
Reč	Šuma	Mama	Baba
Prosečna greška	2.08 %	1,67%	2.91%

Kada se za grešku segmentacije uzme stroži uslov od 2 rama analize, ukupna greška je 5.27%, a pojedinačno za reči /mama/, /baba/ i /šuma/ vrednosti se nalaze u tabeli 5.6.

Tabela 5.6 Greška segmentacije GFCC model – stroži uslov

Reč	Šuma	Mama	Baba
Prosečna greška	4.16%	5%	6.66%

Primeri segmentacije reči /šuma/ pomoću GFCC koeficijenata dat je na slici 5.6. Sa (a) je obeležen prikaz talasnog oblika signala na kome su zelenom bojom obeležene ekspertski određene granice, a crvenom granice dobijene predloženim algoritmom, sa (b) je obeležen primer formiranja putanje najmanje težine, a sa (c) spektrogram testnog signala. Iznad prikaza talasnog oblika signala dodatno su istaknute i obeležene ekspertski određene granice fonema i pauza pre i posle reči.



Slika 5.8 Segmentacija reči korišćenem GFCC koeficijenata

5.3 Rezime

U ovom poglavlju razmatran je problem segmentacije govornog signala prilagođen korišćenju kod patološkog govora. Imajući u vidu metodologiju logopedskog testiranja po kojoj se od ispitanika traži da izgovara unapred definisane stimuluse tačno određenim redom, može se smatrati da tada ovaj problem spada u oblast poluatomatske segmentacije i da je moguće formirati modele reči govornog korpusa. Predložena su dva algoritma segmentacije reči na foneme. Prvi je baziran na algoritmu maksimalne verodostojnosti a drugi na dinamičkom vremenskom usklađivanju. Na osnovu rezultata dobijenih predloženim algoritmima za segmentaciju, može se zaključiti da je prosečna greška segmentacije koja se dobija upotrebom metode bazirane na algoritmu maksimalne verodostojnosti 25% dok je upotrebom algoritma vremenskog usklađivanja ta greška u najstrožem slučaju 5.27%. Imajući to u vidu, u okviru sistema je usvojeno korišćenje algoritma dinamičkog vremenskog usklađivanja za segmentaciju. Takođe, kako je u tom slučaju parametrizacija MFCC koeficijentima rezultirala prosečnim greškama od 0.88%, odnosno od 2.26% a upotreba GFCC koeficijenata prosečnim greškama segmentacije od 2.22%, odnosno 5.27% pod blažim i strožijim uslovima, respektivno, MFCC koeficijenti su usvojeni kao metod izdvajanja obeležja govornog signala. Razlog za ovo možemo tražiti u činjenici da je radi optimizacije brzine obrade podataka korišćen *Lyon*–ov model *gammatone* banke filtara koji ima malu selektivnost na niskim frekvencijama. Pored toga, kada se želi postići velika tačnost segmentacije koristi se kombinacija audio i vizuelnih metoda. Tačnije, govor se istovremeno preslušava i prate se njegov talasni oblik i spektralni sadržaj, jer nije moguće precizno utvrditi granice segmenata oslanjajući se samo na čulo sluha. Imajući to u vidu može se reči da iako se *gammatone* bankom filtara bolje modeluje ponašanje auditornog sistema, ono sa sobom nosi i probleme koji su karakteristični za čulo sluha.

6 Detekcija patologije govora na bazi globalne ocene

Globalni artikulacioni test koristi se za procenu nivoa dostignute razvijenosti i kvaliteta izgovora glasova srpskog jezika. Njime se utvrđuje broj i vrsta oštećenjih glasova kao i stepen oštećenja. Osnovni cilj ovog dela istraživanja je primena neuralnih mreža za detekciju i ocenu poremećaja govora na nivou fonema baziranih na Globalnom artikulacionom testu. One su primenjene jer za prepoznavanje fonema kao osnovnih jedinica govora postoje dobro razvijeni algoritmi bazirani na neuralnim mrežama (Waibel i sar., 1990; Watrous i sar., 1990; Altosaar i sar.,1992) pa se očekuje da se slične metode mogu primeniti i u slučaju prepoznavanja patologije govora.

Veštačke neuralne mreže, oponašanjem atributa ljudskog mozga, pre svega paralelne obrade informacija, prevazilaze mogućnosti konvencionalnih, računarskih procedura za sekvencijalnu obradu informacija. Osnovnu računarsku snagu neuralnih mreža čini masivni paralelizam, sposobnost obučavanja i generalizacija, koja predstavlja sposobnost produkovanja zadovoljavajućeg izlaza neuralne mreže i za ulaze koji nisu bili prisutni u toku obučavanja. Neuralne mreže se odlikuju mogućnošću formiranja znanja na osnovu klasa uzoraka i prepoznavanja na osnovu sličnosti sa obučavajućim uzorcima, pokazujući jako dobra diskriminativna svojstva između klasa. Posebno su efikasne u slučajevima kada u višedimenzionom polju karakteristika treba odrediti pripadnost uzorka jednoj od klasa čak i kada su klase teško separabilne. Zahvaljujući nelinearnosti svojih računarskih jedinica (neurona) dovoljno složena neuralna mreža može biti dobra aproksimacija ma kog nelinearnog dinamičkog procesa, pa i govora. Ova karakteristika neuralnih mreža se posebno ističe kada je u pitanju detekcija obeležja koja utiču na kvalitet izgovornog glasa jer je potrebno razdvojiti klase normalnog i patološkog izgovora posmatrajući realizacije istog fonema.

Efikasnost neuralnih mreža zavisi od njihove topologije i izbora ulaznih parametara pa je neophodno ispitati različite opcije modela predloženog sistema za ocenu kvaliteta govora i uporediti ih u smislu tačnosti, robustnosti i praktične primene. U nastavku su dati primeri detekcije patologije izgovora fonema upotrebom neuralnih mreža, bazirane na globalnom artikulacionom testu.

6.1 Primena neuralnih mreža kod detekcije patološkog govora - pregled literature

Neuralne mreže su našle svoju primenu kod detekcije patologije u govoru i to u koraku gde se vrši razlikovanje patološkog i normalnog govora. Brojni primeri iz literature svedoče da se neuralne mreže različitih oblika mogu uspešno koristiti za detekciju širokog opsega patologija koje se manifestuju kroz degradaciju kvaliteta govora. Tako je u referenci Linder i sar. (2008) neuralna mreža obučavana za prepoznavanje stepena disfonije (bez disfonije, nizak, blag i visok stepen disfonije), a rezultati testiranja na 120 pacijenata koji pate od laringealnih oboljenja pokazala su da se sa pouzdanošću od 80% može odrediti postojanje disfonije. Iz govornog signala izdvojeni su *jitter*, *shimmer*, GNE i varijacije dužine periode osnovnog glasa, dok je za detekciju stepena disfonije korišćen ansambl neuralnih mreža.

Neuralne mreže su sa velikim uspehom korišćene i kod detekcije patološkog govora koji nastaje kao posledica laringealnih oboljenja i neuroloških poremećaja (Parkinsoniva bolest, Alchajmerova bolest, disleksija i sl.). U referenci Salhi i sar. (2008) kao obeležje govornog signala usvojeno je 5 *wavelet* koeficijenata, a neuralna mreža je imala 5 neurana u prvom, 15 u skrivenom i jedan u izlaznom sloju. Na uzorku od 100 reči (50 normalnih i 50 patoloških uzoraka) od kojih je 80 korišćeno za obuku a 20 za testiranje, postignuta je tačnost detekcije patološkog izgovora od 100%. *Wavelet* transformacija u kombinaciji sa neuralnim mrežama je korišćena i za prepoznavanje karakterističnih govornih patologija (Akbari i sar., 2014) koje nastaju usled vokalne hiperfunkcije, disfonije i laringofaringealnog refluksa. Tačnost klasifikacije bila je 97,33% a neuralna mreža je imala 2 skrivena sloja sa 15 i 20 neurona.

MFCC koeficijenti su, tradicionalno, najčešće korišćeni kao način izdvajanja obeležja govornog signala, pa su tako našli primenu i kod neuralnih mreža u svrhu detekcije patološkog govora. Godino-Llorente i sar., (2009) koristili su, pored 14 MFCC koeficijenata, energiju u prozoru analize, tri mere šuma (HNR, NNE i GNE) i prve izvode ovih vrednosti. Neuralna mreža imala je 36 neurona u ulaznom, 100 neurona u skrivenom sloju i jednim na izlazu. Testiranje obavljeno na grupi od 117 patoloških i 23 normalna izgovora, pokazalo je rezultate prepoznavanja patologije od 93.8% za kontinualno izgovaran vokal /a/, i 96.3% u kontinualnom govoru.

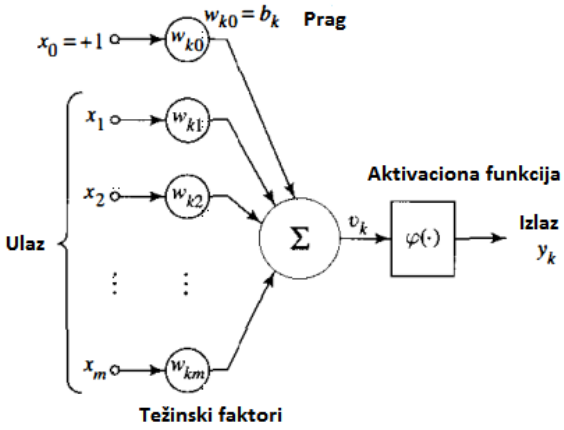
Što se tiče primene vezane za istraživanje artikulacije, neuralne mreže su korišćene za detekciju mesta artikulacije (Paulraj i sar., 2008), kao u okviru sistema za procenu kvaliteta artikulacije. U Imam i sar. 2008, opisan je softver razvijen u okviru programskog paketa *Praat* kod koga je neuralna mreža korišćena za procenu tipa distorzije glasa /r/ u okviru reči. Softver koji je za tu namenu razvijen, baziran je na *feed forward* neuralnoj mreži treniranoj na rečima koje je izgovaralo 70 govornika različitog pola i uzrasta. Glas od značaja (glas /r/) nalazio se u inicijalnoj, medijalnoj i finalnoj poziciji u okviru reči tako da su obuhvaćeni svi slučajevi u kojima se distorzija može javiti. Kao obeležja ispitivanih reči usvojeno je 16 LPC koeficijenata. Neuralna mreža je obučavana *back propagation* algoritmom. Optimizujući topologiju neuralne mreže, autori su postigli najbolju klasifikaciju u slučaju kada je u ulaznom sloju bilo 16 neurona, u prvom i drugom skrivenom sloju po 31 neuron i u izlaznom sloju 5 neurona. Greška klasifikacije je iznosila 25%.

Korišćenje neuralnih mreža za klasifikaciju poremećaja artikulacije tipa substitucije prikazano je u Georgoulas i sar. (2009). *Empirical Mode Decomposition* i *Hilbert Huang* transformacija su korišćeni za formiranje spektralne reprezentacije govornog signala nakon čega su izdvajana obeležja govornog signala u formi mel kepstralnih koeficijenata koja su dovedena na ulaz neuralne mreže. Ispitivanje na grupi od 144-oro dece sa poremećajem artikulacije u formi substitucije glasa /s/, pokazalo je da se ovim postupkom može postići tačnost klasifikacije od 79.86%

6.2 Neuralne mreže – osnovni pojmovi

Neuralne mreže kao pojednostavljen matematički model biološkog nervnog sistema, sastoje se od velikog broja gusto povezanih procesorskih elemenata i vrše distribuiranu paralelnu obradu podataka. Neuralne mreže imaju sposobnost skladištenja znanja stečenog kroz proces učenja u okviru svojih interkonekcija, kao i korišćenja tog znanja (Haykin, 1994). Elementi za procesiranje dinamički odgovaraju na ulazni stimulus, i pri tome je ovaj odgovor kompletno zavisan od lokalne informacije sadržane u njihovom okruženju (Nigrin, 1993).

Osnovni gradivni element neuralnih mreža je neuron. Na slici 6.1 dat je model neurona.

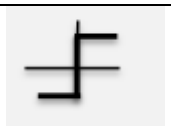
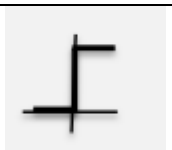
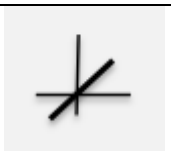


Slika 6.1 Model neurona

Prvi, ulazni deo neurona, prima ulazne podatke i posredstvom mrežne funkcije i težinskih koeficijenata formira vrednost koja se dalje prosleđuje na ulaz drugog dela neurona koji predstavlja aktivacionu funkciju. Mrežna funkcija ima zadatak da kombinuje ulazne veličine i najčešće je u formi linearne kombinacije ulaznih veličina sa težinskim faktorima w. Veličinom praga b_k modeluje se "lokalno polje", odnosno uticaj samog neurona na ulaznu veličinu. Model neurona sa pragom različitim od nule i linearnom aktivacionom funkcijom može se pojednostaviti ukoliko pored postojećih N ulaza uvedemo i N+1 ulaz za koji će x_0 imati vrednost 1. Aktivaciona funkcija u zavisnosti od sume ponderisanih ulaza formira izlaz koji je obično u tačno određenim

granicama tako da većina aktivacionih funkcija za kodomen ima konačan podinterval skupa realnih brojeva: (0, 1) ili (−1, 1). Neke od aktivacionoh funkcija navedene su u tabeli 6.1.

Tabela 6.1 Primeri aktivacionih funkcija

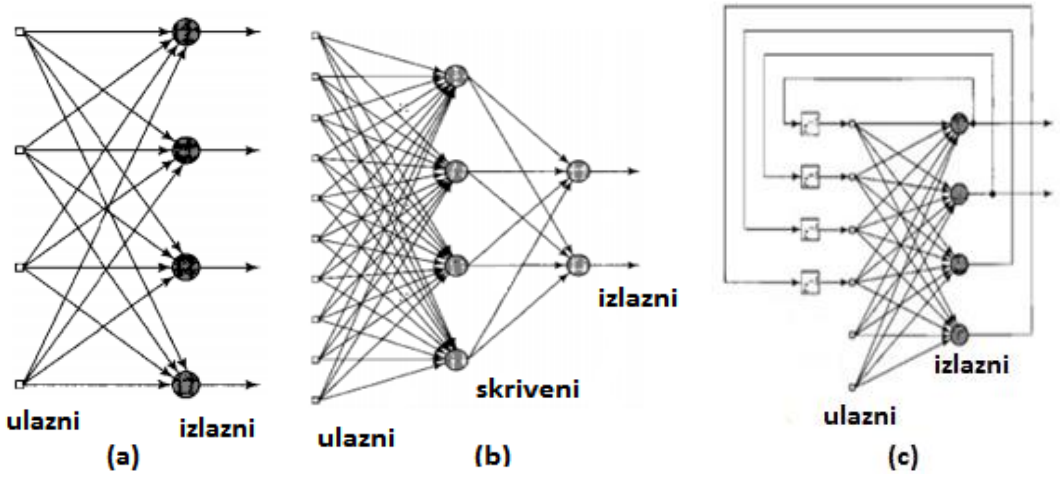
Naziv funkcije	Formula	Grafički prikaz
Funkcija praga - bipolarna	$y(x)=\begin{cases}1, & x>0\\ -1, & x<0\end{cases}$	
Funkcija praga - unipolarna	$y(x)=\begin{cases}1, & x>0\\ 0, & x<0\end{cases}$	
Linearna u delovima	$y(x)=\begin{cases}1 & x>1\\ x & -1<x<1\\ 1- & x<-1\end{cases}$	
Unipolarna sigmoid funkcija	$y(x)=\frac{1}{1+e^{-x}}$	
Bipolarna sigmoid funkcija	$y(x)=\frac{2}{1+e^{-2x}}-1$	

Način na koji su neuroni u neuralnoj mreži struktuirani utiče na njene performanse i način obučavanja. Sa stanovišta ahitekture neuralne mreže generalno se izdvajaju tri klase neuralnih mreža (Haykin, 1994):

1. Jednoslojne *feedforward* mreže, koje predstavljaju najjednostavniji oblik neuralnih mreža koje se sastoje od ulaznog sloja neurona direktno povezanim sa izlaznim slojem, bez povratnih sprega. Na slici 6.2a, dat je primer jednoslojne neuralne mreže koja ima četiri neurona u ulaznom i četiri neurona u izlaznom sloju, naziv "jednoslojni" odnosi se na izlazni sloj jer se u ulaznom sloju ne vrši procesiranje.
2. Višeslojne *feedforward* mreže koje se sastoje od ulaznog sloja, više skrivenih slojeva i izlaznog sloja. Kod njih takođe nema povratnih sprega

između neurona. Ulazni sloj ima istu funkciju kao i kod jednoslojnih *feedforward* mreža, a obrada podataka se vrši u jednom ili više skrivenih slojeva. Pri tome se podrazumeva se da su neuroni prvog skrivenog sloja povezani samo sa neuronima drugog skrivenog sloja i tako redom. Na slici 6.2b dat je primer višeslojne *feedforward* mreže.

3. Rekurentne mreže, kod kojih postoje povratne veze među neuronima, kao što je prikazano na slici 6.2c, gde je izlaz neurona doveden na ulaze drugih neurona. Rekurentne mreže mogu imati i više skrivenih slojeva neurona.



Slika 6.2 Topologije neuralnih mreža

Najbitnija osobina neuralnih mreža jeste njihova sposobnost da stiču znanje putem učenja. One stiču znanje kroz iterativni proces podešavajći težinske faktore sinapsi i njihove pragove. Prema vrsti obuke neuralne mreže, razlikuju se:

1. Obučavanje sa nadgledanjem (*Supervised learning*) kod koga je prisutan obučavajući skup u formi parova ulaznih parametara i odgovarajućih izlaznih parametara.

2. Obučavanje sa podsticanjem (*Reinforcement learning*) gde neuralna mreža dobija rudimentirane informacije o tome kakav izlaz produkuje, najčešće samo u formi jednog bita informacije tipa dobar/loš.

3. Samoobučavanje (*Unsupervised learning*) je karakterisano odsustvom bilo kakve informacije od strane okruženja po pitanju izlaznih parametara mreže.

Jedna od najpopularnijih struktura neuralnih mreža svakako je višeslojna *feedforvard* neuralna mreža, ili, alternativno, višeslojni perceptron (*multilayer perceptron* - MLP). Višeslojni perceptron uspešno je korišćen za rešavanje složenih problema zahvaljujući algoritmu obučavanja poznatog pod nazivom - algoritam propagacije greške unazad (*back propagation algorithm*). *Back propagation* algoritam spada u algoritme obučavanja sa nadgledanjam i generalno gledajući, izvodi se putem dva prolaza kroz neuralnu mrežu: prolaza unapred (u pravcu prostiranja sinaptičkih veza) i prolaza unazad (suprotno od pravca prostiranja sinaptičkih veza). U prvom prolazu, za vrednosti dovedene na ulaz neuralne mreže, formira se odgovarajući izlaz. Pri tome, težinski faktori sinapsi ostaju fiksirani. Na osnovu dobijene izlazne vrednosti i željene (ciljane) vrednosti izlaza, izračunava se vrednost greške. U prolazu unazad, vrednost greške koristi se kao parametar na osnovu koga se koriguju težinski faktori sinapsi. Ovaj proces se ponavlja za ceo set parova ulaznih i izlaznih vrednosti. Jedan prolazak algoritma kroz ceo obučavajući skup predstavlja epohu. Obučavanje se vrši kroz više epoha sve dok se ne postigne željena tolerancija greške, odnosno sve dok se težinski faktori sinaptičkih veza i pragova ne ustale i srednja kvadratna greška za ceo obučavajući skup ne konvergira ka nekoj minimalnoj vrednosti. Velikoj mreži može biti potrebno dosta vremena za konvergenciju, tako da se ponekad umesto veličine greške zadaje broj epoha. Kako se znanje neuralne mreže sadržano u težinskim faktorima sinapsi formira na osnovu parova ulaznih i izlaznih veličina koje učestvuju u obuci, neuralna mreža se obučava da ispoljava najmanju grešku za obučavajući skup. Jedan od problema koji se javlja kod neuralnih mreža jeste *overfitting*, odnosno, pojave da neuralna mreža za obučavajuči skup ima vrlo malu grešku, ali za nove ulazne podatke daje veliku grešku. Drugim rečima, neuralna mreža je memorisala obučavajući skup i nije naučila da generalizuje.

6.3 Detekcija artikulacionih poremećaja pomoću neuralnih mreža

Istraživanja na polju detekcije globalnih poremećaja artikulacije primenom neuralnih mreža podrazumevala su postepeno usložnjavanje topologije mreža i optimizaciju ulaznih parametara. U ovom delu, najpre su prikazani najbitniji elementi neuralnih mreža, nakon čega su dati rezultati istraživanja na polju klasifikacije normalnog i patološkog govora pomoću višeslojnog perceptrona (VP) i ansambla neuralnih mreža.

6.3.1 Detekcija artikulacionih poremećaja pomoću višeslojnog perceptrona

Prva istraživanja na temu klasifikacije artikulacionih poremećaja glasova srpskog jezika pomoću neuralnih mreža izvedena su korišćenjem višeslojnog perceptrona (Furundžić i sar. 2006, 2007, Bilibajkic i sar., 2014). Neuralna mreža sastojala se od ulaznog, jednog skrivenog sloja i izlaznog sloja.

U ovom slučaju, skup od 194 izgovora foneme /š/ je korišćen za klasifikaciju normalnog i patološkog govora. Ulazni vektor podataka od 12 MFCC koeficijenata proširen je sa 4 parametra koja se odnose na talasni oblik i energiju govornog signala. Posmatrano je trajanje reči, trajanje fonema i energija reči i energija fonema, što, pored navedenih MFCC parametara čini vektor od 16 ulaznih parametara. Skup izgovornih fonema podeljen je na dva jednaka dela, obučavajući i testni uzorak. Optimalan broj neurona u skrivenom sloju je bio 15, dok je izlazni sloj činio jedan neuron. Za obučavanje je korišćen *Levenberg-Marquardt backpropagation* algoritam sa adaptivnim momentom. Sprečavanje o*verfitting-a* izvedeno je postavljanjem odgovarajućeg seta validacionih parametara. Klasifikacija na obučavajućem skupu pokazala je 100% tačnost dok je na testnom skupu taj procenat bio 96%.

Pored globalne ocene poremećaja artikulacije, posmatrane su i ocene sa analitičkog testa kao ciljne vrednosti. Testiranje je izvedeno na grupi od 200 dece, od

toga sto dečaka i sto devojčica. Uzorak je podeljena na dva dela i to na testni i obučavajući skup. Distribucija ispitanika prema kvalitetu izgovora glasova /š/ i /ž/ je bila takva da je 50% ispitanika bilo sa korektnim izgovorom, što dalje znači da je skup izbalansiran. Evaluacija kvaliteta izgovora vršena je globalnim artikulacionim testom i analitičkim testom od strane osam logopeda, a konačna ocena je formirana kao srednja vrednosti pojedinačnih vrednosti. Obučavajući skup sastojao se od 100 ulaznih vektora koji predstavljaju karakteristike govornih segmenata i odgovarajućih 100 vektora binarnih vrednosti (0, 1) koji označavaju korektan i patološki izgovor. Pod korektnim izgovorom podrazumevaju se ocene 3 i 4 na globalnom artikulacionom testu, dok su ocene 5 i 6 smatrane nekorektnim izgovorom. Kod analitičkog testa, prisustvo bilo koje devijacije smatrano je nekorektnim izgovorom. Optimalna struktura višeslojnog perceptrona dobijena je tako što je postepeno povećevana kompleksnost skrivenog sloja. Najbolji rezultati postignuti su mrežom koja je sadržala 10 skrivenih i jedan izlazni sloj. Ulazni vektor je bio dimenzije 17, i sastojao se od 12 MFCC parametara, dužine reči, dužine inicijalne foneme, njihovog odnosa i srednje vrednosti i standardne devijacije energije inicijalnog fonema. *Levenberg-Marquardt* algoritam sa adaptivnim momentom je korišćen za obuku kako bi se povećala efikasnost obučavanja. I u ovom slučaju je *overfitting* kontrolisan skupom validacionih parametara. Performanse modela date su u tabeli 6.2.

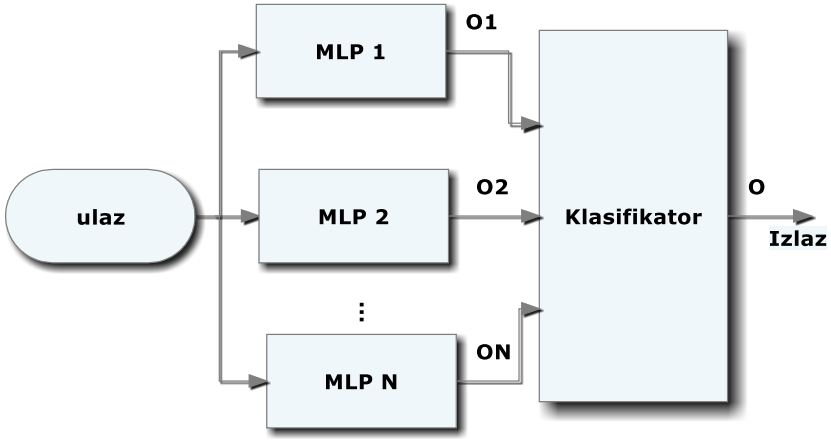
Tabela 6.2 Rezultati prepoznavanja patološkog izgovora pomoću višeslojnog perceptrona

Model/metoda	Tačnost, Obučavajući skup [%]		Tačnost, Testni skup [%]	
	/š/	/ž/	/š/	/ž/
GAT/MFCC	100	100	96.66	96.55
AT/MFCC	82.66	79.28	76.41	78.35

6.3.2 Primena ansambla neuralnih mreža kod detekcije artikulacionih poremećaja

Dalji razvoj modela postignut je unapređenjem metode klasifikacije uvođenjem ansambla neuralnih mreža (Furunžić i sar., 2013). Pod ansamblom neuralnih mreža

smatra se model koji se sastoji od skupa neuralnih mreža – "eksperata" koje su obučene da izvršavaju isti zadatak. Svaka neuralna mreža je nezavisno trenirana istim skupom podataka, a ansambl objedinjava znanje pojedinačnih eksperata i donosi odluku koja je superiorna u odnosu na onu donesenu od strane pojedinačnog eksperta, kao što je šematski prikazano na slici 6.3. Kod korišćenja ansambla eksperata, realno je očekivati da pojedinačne neuralne mreže teže različitim lokalnim minimumima, pa se tako kombinacijom performansi poboljšava ukupan rezultat (Sollich i Krogh, 1990; Hansen i Salamon, 1996). Glavne prednosti ovakvog pristupa ogledaju se u poboljšanoj moći generalizacije, većoj robusnosti i smanjenoj kompleksnosti strukture.



Slika 6.3 Ansambl neuralnih mreža

Algoritam formiranja ansambla izveden je pomoću sledećih koraka:

1. Izdvojena su dva podskupa iste veličine, obučavajući i testni, vodeći računa o jednakom prisustvu obe klase izgovora.
2. Izbalansirana je potencijalna razlika između klasa.
3. Uzeto je 80% obučavajućeg skupa za obuku a preostalih 20% za validaciju.
4. Određena je struktura višeslojnog perceptrona postepeno povećavajući kompleksnost strukture neuralne mreže.
5. Više puta je ponovljena procedura treniranja i dobijeni rezultati su sačuvani u formi vektora težina i pragova.
6. Korišćenjem testnog uzorka određen je izlaz za svaki višeslojni perceptron koji predstavlja eksperta.

7. Definisane su performanse eksperata kao srednja kvadratnu grešku ocene. U ovom slučaju korišćen je ansambl od 100 eksperata.

8. Eksperti su sortirani prema performansama u rastućem redosledu

9. Uzeto je M najbolje treniranih eksperata, u našem slučaju to je bilo M=20. Verovatnoća greške pojedinačnog eksperta nije prelazila 0.5.

10. Uzet je podskup od N eksperata iz skupa od M, formirano je svih C kombinacija od N elemenata iz skupa od M elemenata.

$$C = \frac{M!}{N!(M-N)!} \tag{6.1}$$

U našem slučaju poredili smo rezultate dobijene za ansambl od $N \in \{4,5,6,10\}$

11. Za svaku od C kombinacija N eksperata izračunata je vrednost O_c koja predstavlja izlaznu vrednost za ceo ansambl i računa se kao:

$O_c = \frac{1}{N}\sum_{i=1}^{N} O_i$, gde su sa O_i označene pojedinačne vrednosti izlaza svakog od "eksperata" iz ansambla.

12. Izračunata je greška za svaki ansambl kao $E_c = mse(t-O_c)$ gde je *mse* srednja kvadratna greška a c = 1,2,… C.

13. Određen je najbolji ansambl kao $\arg\min_c (E_c)$

Za testiranje je uzeto 200 uzoraka glasova /š/ i /ž/ iz reči /šuma/ i /žaba/ iz baze dečijeg govora opisane u poglavlju 4.5. Uzorci su ocenjeni globalnim artikulacionim testom i analitičkim testom. Svakom od uzoraka je pridružena binarna vrednost (0 ili 1) koja označava kvalitet izgovorenog glasa na osnovu ocena globalnog artikulacionig testa. Svaki od uzoraka je parametrizovan tako što je formiran vektor od 17 vrednosti koje su činili 13 MFCC koeficijenata, dužina trajanja govornog stimulusa, standardna devijacija i srednja vrednost energije, i energija sadržana u inicijalnom fonemu /š/, odnosno /ž/ i energija susednog vokala /u/ odnosno /a/. Višeslojni perceptron je imao, pored ulaznog, jedan skriveni sloj i jedan neuron u izlaznom sloju. Obuka je vršena pomoću *Levenberg-Marquardt* algoritma. Pojava *overfitting*-a je kontrolisana skupom validacionih parametara. Rezultati su prikazani u Tabeli 6.3.

Tabela 6.3 Ansambl neuralnih mreža kod prepoznavanja patologije govora

Model	Tačnost [%]	
	Obučavajući skup	Testni skup
Ansambl 4	100.00	85.00
Ansambl 5	100.00	88.00
Ansambl 6	100.00	91.00
Ansambl 10	100.00	98.00

6.4 Rezime

U ovom poglavlju prestavljan deo sistema za prepoznavanje patološkog izgovora čija je funkcija formiranje globalne ocene artikulacije izgovorenog glasa. Ovaj modul sistema trebao bi da automatizuje ocenjivanje koje se u logopedskoj praksi obavlja pomoću globalnog artikulacionog testa koji se koristi za procenu nivoa dostignute razvijenosti i kvaliteta izgovora glasova srpskog jezika. U tu svrhu korišćene su veštačke neuralne mreže. Kroz pregled literature može se zaključiti da su neuralne mreže dale dobre rezultate na polju prepoznavanja patološkog govora, zahvaljujući svojim osobinama kao što su sposobnost generalizacije i učenja na ograničenom broju uzoraka. Istraživanja na temu globalne ocene patologije govora vršena su pomoću dva oblika neuralnih mreža: višeslojnog perceptrona i ansambla neuralnih mreža. Rezultati testiranja na govornoj bazi dečijeg govora pokazali su da se pomoću višeslojnog perceptrona i MFCC parametara kao obeležja govornog signala postiže prosečna tačnost od 96,6% za razdvajanje klasa normalnog i patološkog govora. Korišćenjem ansambla neuralnih mreža naveća tačnost razdvajanja postiže se za slučaj kada u ansamblu ima 10 "eksperata", i ona iznosi 98%. Zbog boljih performansi po pitanju tačnosti prepoznavnja patologije, ansambl neuralnih mreža je prihvaćen kao deo sistema kojim se rešava problem detekcije globalne ocene artikulacije.

7 Detekcija patologije govora na bazi analitičke ocene

Imajući u vidu raznovrsnost ispoljavanja devijacija u govoru, kao i činjenicu da se pojedine pojave u patološkom izgovoru mogu naći samo u okviru određene grupe glasova, algoritamski postupci detekcije patologije u govoru moraju se rešavati na nivou konkretne patološke pojave. Treba imati u vidu i to da je parametre algoritma neophodno podesiti, tj opitimizovati u zavisnosti od analiziranog glasa, a prema vrednostima dobijenim na osnovu eksperimentalne analize normalnog i patološkog izgovora, odnosno kriterijuma diferencijacije tipičnog i atipičnog izgovora.

Da bi bilo moguće algoritamski detektovati devijacije potrebno je odrediti one akustičke parametre pomoću kojih je moguće razdvojiti oblasti normalnog i patološkog izgovora kao i odgovarajuće granične vrednosti u parametarskom domenu. Obimna istraživanja na temu artikulaciono-akustičkih odstupanja patološkog izgovora (Punišić, 2012; Punišić i sar. 2012) omogućila su do izvesne mere definisanje navedenih oblasti. Pored toga, istraživanja na govornoj bazi globalnog artikulacionog testa i analitičkog testa (Punišić, 2012.) pokazala su da se najveća učestalost atipične produkcije tipa distorzije javlja u izgovoru 7 fonema (/c/, /č/, /dž/, /š/, /ž/, /r/ i /l/). U ovoj studiji utvrđeno je da se različite vrste sigmatizama, promenjenog kvaliteta trajanja i izmenjenog intenziteta frikcije pre svega u spektralnom domenu, ističu kao najčešća odstupanja od normalnog izgovora. Upravo iz tih razloga, istraživanja prikazana u ovom poglavlju su usmerena na analizu distorzija tipa trajanja i jedne vrste sigmatizma - stridensa.

7.1 Detekcija stridensa

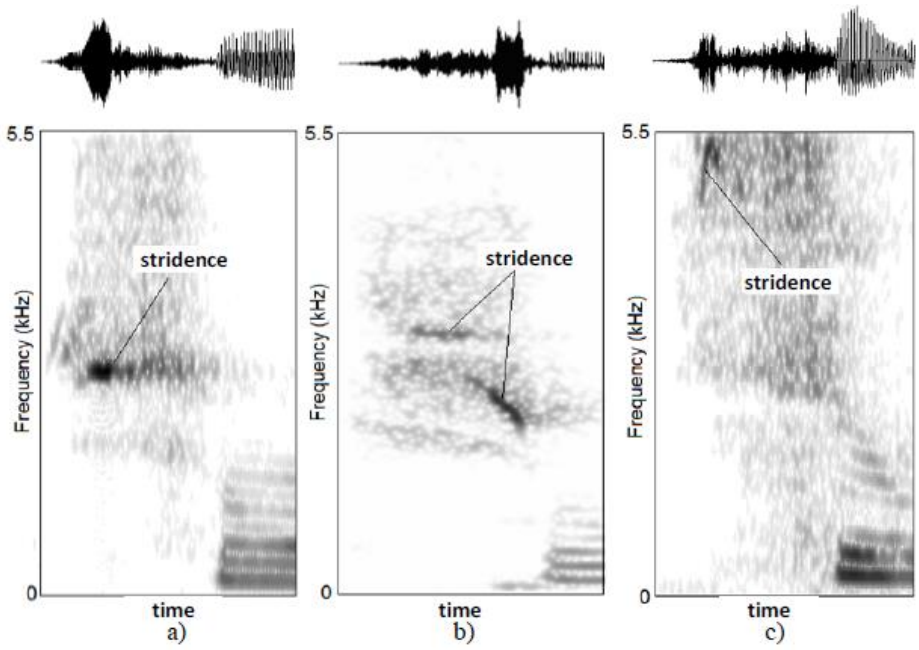
Stridens, odnosno, stridentni sigmatizam predstavlja jednu od atipčnih relizacija glasova srpskog jezika koja se prvenstveno javlja kod izgovora frikativa i afrikata. U perceptivnom domenu doživljava se kao neprijatan zvuk nalik zvižduku ili kao piskavi ili hrapav zvuk koji se javlja uporedo sa normalnim izgovorom, odnosno sa normalnom frikcijom. Ova pojava nastaje kao posledica određenog položaja jezika u odnosu na zube i nepce. Naime, prilikom artikulacije stridens može nastati formiranjem žljeba po sredini jezika koji je pokriven gornjim nepcem, ili formiranjem rezonantne šupljine, između vrha jezika i gornjih sekutića, i slično. Prilikom prolaska kroz ovako postavljene artikulacione organe, vazdušna struja postiže određenu brzinu i dolazi do formiranja stojećih talasa i oscilovanja, pri čemu se generiše ton određene frekvencije ili vrlo snažan uskopojasni rezonantni šum. Ovakva nepravilnost koja se javlja prilikom artikulacije frikativa i afrikata za posledicu ima promenu akustičkih karakteristika frikcije kojom se narušava kvalitet izgovora fonema i smatra se patološkim izgovorom u srpskom jeziku. Pojava stridensa može se sporadično javiti i kod normalne artikulacije, posebno u nekim prelaznim položajima artikulacionih organa prilikom koartikulacije. Takav stridens je sporadična i kratkotrajna pojava i ne smatra se patologijom. Treba napomenuti da u nekim drugim jezicima, na primer u engleskom, stridentnost predstavlja distinktivno obeležje izgovora glasova (Chomsky i Halle, 1968). Takođe, u nekim jezicima zviždući frikativi (stridentni) predstavljaju pravilan izgovor (Shosted, 2006), međutim, u srpskom (Jovičić i sar., 2008) i češkom (Honova i sar., 2003) i još nekim jezicima predstavljaju nepravilan izgovor.

Na slici 7.1 dati su tipični primeri stridensa kod frikativa /š/ koji se nalazi u inicijalnom poziciji u reči /šuma/. Prvi primer (slika 7.1a) ilustruje vremenski stabilnu, intenzivnu i uskopojasnu rezonantnu pojavu usađenu u difuzni šumni spektar frikativa /š/. Na slici 7.2b prikazana je pojava dvostrukog stridensa, odnosno, na višim frekvencijama javlja se vremenski stabilna rezonansa, dok se na nižim frekvencijama javlja rezonantna pojava koja je frekvencijski promenljiva. Treći primer na slici 7.1c ilustruje pojavu kratkotrajnog frekvencijski varijabilnog stridensa.

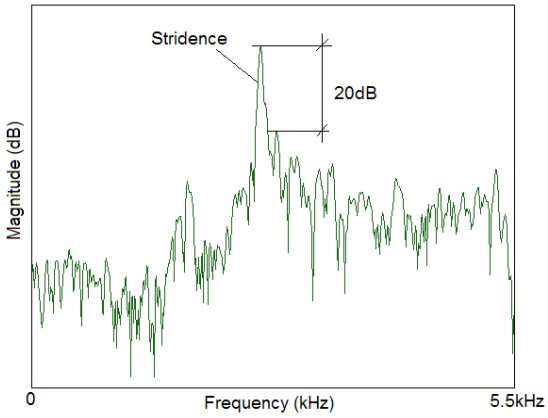
Kratkovremeni spektar frikativa /š/ na srednjem delu stridensa datog na slici 7.1a prikazan je na slici 7.2. Uočava se da je anvelopa signala takva da je intenzitet na

mestu pojave jake rezonanse koja predstavlja stridens 20 dB iznad anvelope signala ostatka spektra. Ovakva frekvencijska karakteristika je jedan od uslova da bi rezonatna pojava u spektru bila percepirana kao stridens. Dodatni uslov je još i da rezonantna pojava ima minimalno trajanje oko 10 ms (Jovičić i sar., 2008). Takođe, posmatrajući spektrograme i poredeći ih sa vremenskom predstavom signala u vidu talasnog oblika uočava se jaka korelacija između talasnog oblika signala i spektrograma (slika 7.1) na mestu pojave stridensa.

Navedene karakteristike stridensa u akustičkom domenu korišćene su kao kriterijumi za algoritamsku detekciju ove vrste atipičnog izgovora.



Slika 7.1 Tipičan primer stridensa u izgovoru glasa /š/ u inicijalnoj poziciji u reči /šuma/



Slika 7.2 Spektar tipičnog stridensa

Imajući u vidu predhodno opisane karakteristike stridensa kao i način njegove percepcije, postoje naznake da je moguće programski rešiti detekciju stridensa i napraviti programski modul koji bi automatski detektovao pojavu stridensa koja bi se poklapala u manjoj ili većoj meri sa ocenama datim od strane treniranih eksperata. Istraživanja u oblasti automatske detekcije i ocene stridensa data u (Bilibajkić i sar., 2015; Bilibajkić i sar., 2013; Bilibajkić i sar., 2012; Jovičić i sar., 2008) zasnovana su na pronalaženju efikasnog algoritma za detekciju stridensa imajući u vidu njegovu manifestaciju u vremensko-frekvencijskom domenu (Punišić i sar., 2012).

U nastavku su opisane tri metode za automatsku detekciju stridensa. Prve dve prikazane metode se baziraju na modelovanju procesa percepcije govora. Postupak zasnovan na auditornom modelu na bazi *gammatone* banke filtara dat je u poglavlju 7.1.1. Odabrani model dobro "imitira" procese percepcije zvuka u ljudskom auditornom sistemu omogućavajući primenu svih značajnijih psihoakustičkih efekata od kojih je za percepciju stridensa najznačajniji efekat maskiranja. Ovim je postignuta visoka usaglašenost automatske ocene stridensa sa ocenom obučenog profesionalca.

U poglavlju 7.1.2 opisan je postupak koji se bazira na FFT transformaciji i *Burg*-ovom postupku ocene spektra metodom maksimalne entropije (Marple, 1987). Ovim postupkom se relativno pouzdano detektuje stridens u slučajevima kada je naglašen i nesporan, međutim, konstatovano je da u izvesnim slučajevima slabo naglašenog stridensa postoji odstupanje od ocene obučenog ocenjivača.

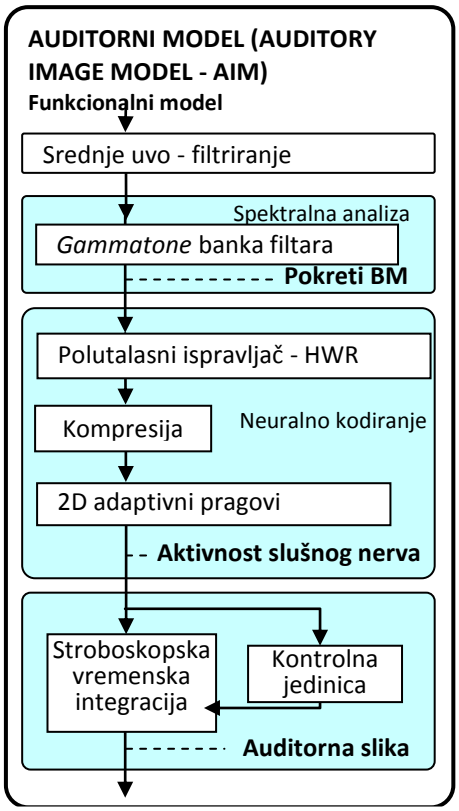
7.1.1 Metod baziran na *Patterson*-ovom auditornom modelu

Istraživanja sprovedena na polju fiziologije, psihologije, a naročito psihoakustike u kombinaciji sa tehnikama obrade signala omogućila su kreiranje matematičkih i računarskih modela koji (pod određenim uslovima i do određenog stepena) mogu da opišu i simuliraju ili "imitiraju" auditorni mehanizam, kao i da daju detaljniji uvid u neke od njegovih elemenata. Dau (2009) je dao pregled auditornih modela koji su generalno podeljeni na biofizičke, psihološke, statističke i perceptivne u zavisnosti od karakteristika na koje se ovi modeli fokusiraju. Treba napomenuti da većina njih nije kompletna, što znači da svaki od njih simulira specifičnu opsevaciju i/ili

funkciju (intenzitet, *pitch*, glasnost i sl.) i, posledično, se može koristiti samo za određeni zadatak. Što se modela auditivne percepcije tiče, razvijeno je više modela kojima se kvantitativno simulira procesiranje signala koje se odvija u kohlei (Hohman, 2002; Slaney i Lyon, 1993; Seneff 1988; Patterson i Allerhand 1995; Patterson i Holdsworth; Lyon, 1984; Shamma,1988; Deng i sar.,1988; Dau i sar., 1997a,b; Jepsen i sar., 2008). Klasični auditorni modeli predloženi 80-tih godina prošlog veka od strane Seneff (1988), Ghitza (1968) i Lion (1982, 1983), korišćeni su kao ulazni moduli kojima se modeluje slušna periferija i bili su vezani prevashodno za sisteme za automatsko prepoznavanje govora (*ASR - automatic speech recognition*). Kako modelovanje fizioloških principa kao takvih u ovom slučaju nije od značaja, već je interesovanje usmereno ka odzivu koji se dobija na određenu pobudu, u ovom slučaju širokopojasnog zvuka kao što su govor i muzika, koriste se tzv. funkcionalni modeli koji simuliraju eksperimentalno utvrđene zakonitosti između ulaza i izlaza auditornog sistema. Kod navedenih funkcionalnih modela odziv bazilarne membrane ne modeluje se kohlearnom hidrodinamikom, već se inicijalna spektralna analiza vrši bankom spektralnih filtara (Patterson i sar., 1987; Irino i Patterson, 1997, 2001; Shekhter i Carney, 1997). Isto tako, funkcionalni model unutrašnje slušne ćelije modeluje se određenom vrstom kompresivne adaptacije kojom se simulira neuralni konvertor. Formiranje auditorne slike kod nekih modela se vrši uz pomoć vremenske integracije (Patterson i Allerhand, 1995; Patterson, 2000) ili izračunavanja korelograma (Slaney i Lyon, 1993; Slaney i sar., 1994).

Rezultati prvih istraživanja na polju automatske detekcije stridensa pomoću auditornih modela data su u Bilibajkić i sar., 2012, gde je korišćena *Lyon*-ova banka filtara, a poboljšanja u Bilibajkić i sar. 2013. U ovom slučaju koristi se *Patterson*-ov auditorni model za formiranje auditorne slike (Patterson i sar., 1995, Patterson i Holdsworth, 1993). Struktura modela prikazana je na slici 7.3. Ovim modelom se transformiše složeni zvuk (pomoću funkcionalnih modela bazilarne membrane i slušnih ćelija), u višekanalnu predstavu strukture neuralne aktivnosti (*neural activity pattern* - NAP) nalik onom koji se opaža kod auditornog nerva, nakon čega biva konvertovan u auditornu sliku koja predstavlja inicijalnu impresiju zvuka. On se sastoji od tri koraka procesiranja.

U prvom koraku vrši se konverzija talasnog oblika signala u odgovarajuću reprezentaciju pokreta bazilarne membrane (*basilar membrane motion* - BMM). Spektralna analiza se izvodi korišćenjem banke auditornih filtara (kojima se ulazni signal konvertuje u niz filtriranih vrednosti). Centralne frekvencije filtara linearno su distribuirane na frekvencijskoj ERB (*equivalent rectangular bandwidths*) skali (Glasberg i Moore, 1990; Patterson i Allerhand, 1995.). Primer kohleagrama koji se dobija ovim filtriranjem prikazan je na slici 7.4a.

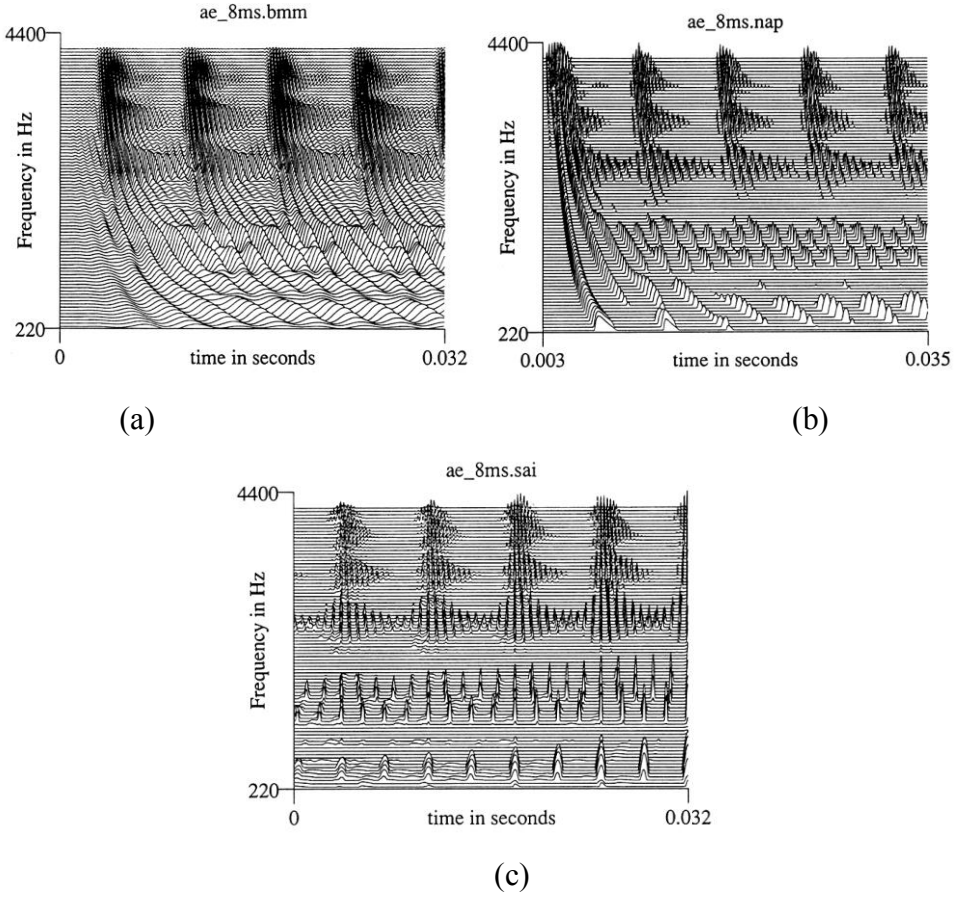


Slika 7.3 Patterson-ov auditorni model

U drugom koraku *Patterson*-ovog modela simulira se pretvaranje mehaničkih pokreta bazilarne membrane u neuralne impulse koji se odigrava putem unutrašnjih slušnih ćelija. Njime se pokreti bazilarne membrane konvertuju u prikaz aktivnosti slušnog nerva (*neural activity pattern* - NAP). NAP se simulira bankom dvodimenzionalnih adaptivnih pragova (Patterson i Allerhand 1995, Patterson i Holdsworth 1996). Mehanizam adaptivnih pragova je reprezentacija neuralnog kodiranja u funkcionalnom modelu. Najpre se vrše ispravljanje (*half wave rectifier*-HWR) i kompresija pokreta BMM, nakon čega se primenjuje vremenska adaptacija i

frekvencijsko potiskivanje. Adaptacija i potiskivanje su upareni i zajednički izoštravaju karakteristike kao što su formanti vokala u BMM reprezentaciji.

U trećem koraku vrši se vremenska integracija pomoću koje se stabilizije repetitivna struktura NAP i dobija se simulirana perceptivna predstava govornog signala koja se naziva auditorna slika (slika 7.4c). NAP se konvertuje u auditornu sliku koristeći banku modula za vremensku integraciju, od kojih svaki modul odgovara jednom kanalu NAP-a. Svaki modul prati aktivnost u kanalu kom je namenjena i vrši stabilizaciju auditorne slike stroboskopskom vremenskom integracijom (s*troboscope time integration* – STI), (Patterson i Allerhand, 1995; Patterson, 2000), koja se kontroliše pomoću lokalnog maksimuma u NAP nizu. Ovakvom vremenskom integracijom NAP periodičnog signala se konvertuje u statičku auditornu sliku dok se za dinamički signal dobija promenljiva slika u kojoj se promene dešavaju onom brzinom kojom se to dešava i u zvuku.



Slika 7.4 Faze analize Patterson-ovog modela
(a) kohleagram, (b) aktivnost slušnog nerva (c) auditorna slika, samoglasnika /ae/ u engleskoj reči /past/ (Patterson i Holdsworth, 1996)

7.1.1.1 Algoritam za detekciju stridensa

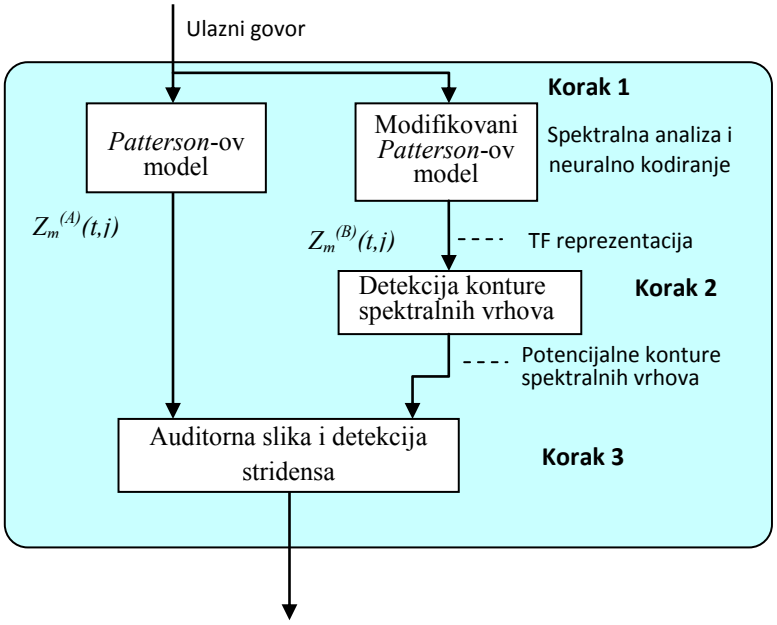
Za procenu mere prisustva stridensa koristi se gore opisani *Patterson*-ov model. Unutar ovog modela "ugrađen" je modul za isticanje spektralnih vrhova i detekciju konture koja potencijalno predstavlja stridens. Mera stridensa se računa na odabranoj konturi spektralnih vrhova koja ima najveću verovatnoću da bude percepirana kao stridens. Za određivanje najbolje konture spektralnih vrhova ne koristi se vremensko-frekvencijska (TF) reprezentacija originalnog *Patterson*-ovog modela već njegova modifikacija u kojoj je polutalasni ispravljač kanalnih signala HWR zamenjen računanjem modula kompleksnog broja.

U odnosu na orginalni *Patterson*-ov model uneta je još jedna izmena. Izostavljena je adaptacija u vremenu. Ovo je učinjeno jer se detekcija stridensa primenjuje na izolovanim fonemima, frikativima, čije je trajanje relativno kratko u odnosu na vreme relaksacije adaptivnog praga.

Na slici 7.5 prikazana su tri koraka kroz koji se realizuje detekcija stridensa. U koraku 1 računaju se dve vremensko-frekvencijske reprezentacije signala u dve paralelne grane obrade, od kojih jedna predstavlja *Patterson*-ov model, a druga modifikovani *Patterson*-ov model, sa izlaznim signalima $z_m^{(A)}(t,j)$ i $z_m^{(B)}(t,j)$, redom.

U koraku 2 na signalu $z_m^{(B)}(t,j)$ lociraju se konture spektralnih vrhova koje mogu biti identifikovane kao stridens. Među izdvojenim konturama se bira ona koja ima najveću verovatnoću da predstavlja stridens.

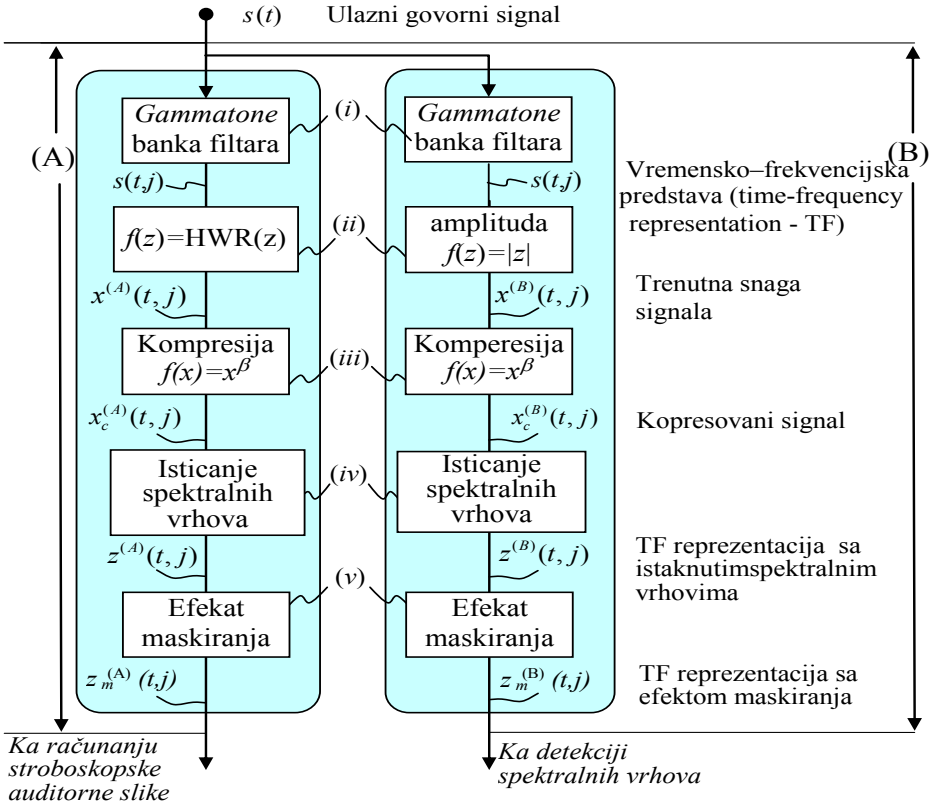
Provera da li izabrana kontura zaista predstavlja stridens realizuje se u koraku 3. Koristeći TF predstavu $z_m^{(A)}(t,j)$ iz prvog koraka obrade, na odabranoj konturi iz koraka 2 formira se auditorna slika STI postupkom opisanim u (Patterson, 2000; Patterson i Holdsworth, 1996). Na osnovu tako dobijene auditorne slike računa se mera subjektivnog osećaja stridensa i donosi odluka o prisustvu stridensa. U sledećim pododeljcima biće detaljnije opisani navedeni algoritamski koraci.



Slika 7.5 Blok dijagram detekcije stridensa

7.1.1.2 Spektralna analiza i neuralno kodiranje

Faze obrade u koraku 1 prikazane su na slici 7.6. U grani A koja se odnosi na *Patterson*-ov auditorni model, prva faza obrade (*i*) je filtriranje *gammatone* bankom filtara GFB kojom se modelira akustički prenos koji odgovara pojedinim tačkama na bazilarnoj membrani. Broj kanala banke filtara *Nch* se bira tako da se postigne potrebna frekvencijska rezolucija, uz prihvatljivo računarsko opterećenje procesora. U zavisnosti od raspoloživih računarskih resursa broj kanala može biti od nekoliko desetina pa do nekoliko stotina. Centralne učestanosti i širine propusnih opsega kanala su raspoređene prema ERB (Hohmann, 2002). U ovom radu je *gammatone* banka filtara realizovana pomoću kompleksnih filtara četvrtog reda (Hohmann, 2002). Razlog primene kompleksnih filtara je mogućnost preciznijeg određivanja anvelope kanalnih signala što je važno za nalaženje kontura spektralnih vrhova. Višekanalni signal na izlazu iz GFB označićemo sa $s(t,j)$, gde je $j = 1,...,N_{ch}$ indeks kanala.



Slika 7.6 Spektralna analiza i neuralno kodiranje
(A) Patterson-ov model, (B) modifikovani Patterson-ov model.

U drugoj fazi obrade (*ii*) u grani (A) primenjuje se polutalasni ispravljač HWR koji modelira procese na trepljastim ćelijama receptora. Polutalasni ispravljač se primenjuje na realni deo kompleksnog signala *s(t,j)*,

$$x^{(A)}(t,j) = HWR\big(s(t,j)\big), \quad HWR(z) = \max\big(real(z),0\big) \tag{7.1}$$

gde je *t* vremenski indeks, a *j* kanalni indeks. Izlaz iz HWR modula je povorka impulsa $x^{(A)}(t,j)$ u kojoj je pored informacije o trenutnoj snazi signala sadržana i informacija o faznom stavu kanalnog signala. Početak svakog impulsa odgovara nultoj fazi kanalnog signala. Stabilnost periode ponavljanja impulsa utiče na oblik signala u auditornom baferu.

Amplitudska kompresija data je na slici 7.6, faza (*iii*) i realizovana je prema,

$$x_c^{(A)}(t,j) = f\big(x^{(A)}(t,j)\big), \;\; f(x) = x^{\beta} \tag{7.2}$$

gde je β pozitivna konstanta kojom se definiše stepen kompresije (Feldbauer i sar., 2005; Patterson and Holdsworth, 1996). Tipična vrednost za β je 0.4 (Feldbauer i sar.,

2005). Ova faza je slična logaritamskoj kompresiji amplitude koja se koristi kod talasnih kodera (npr, μ-law).

Informacija o prisustvu stridensa je sadržana u spektralnim vrhovima. Iz tog razloga se u modulu (*iv*) primenjuje isticanje spektralnih vrhova na sličan način kao i kod algoritama (Patterson i Holdsworth, 1996; Taplidou i Hadjileontiadis, 2007). Isticanje spektralnih vrhova se obavlja u dva koraka. U prvom koraku se vrši filtriranje u frekvencijskom domenu *moving average* (MA) filtrom širine 2*L*+1 tačaka,

$$y(t,j)=\frac{1}{2L+1}\sum_{l=j-L}^{j+L} w(l-j+L+1)x_c^{(A)}(t,l)\,,\quad j=L+1,...,Nch-L \tag{7.3}$$

gde su sa *w*(*i*), *i*=1,...,2*L*+1 obeleženi koeficijenti MA filtra, a sa *y(t,j)* izlaz MA filtra. U sledećem koraku, računa se razlika između elemenata nizova $x_c^{(A)}(t,j)$ i *y(t,j)*. Negativna razlika na konveksnim delovima je postavljena na nulu pomoću

$$z^{(A)}(t,j)=\max\left\{x_c^{(A)}(t,j)-y(t,j),0\right\}. \tag{7.4}$$

Rezonantni ton koji se perceptivno identifikuje kao stridens može da bude maskiran od strane ostalih spektralnih komponenti posebno u slučaju frikativa sa jakom frikcijom. Kao posledica prirodnog procesa maskiranja može da izostane percepcija stridensa. Modeliranje efekta maskiranja se realizuje u modulu (*v*). Spektralne komponente u Δ okolini rezonantne učestanosti *f*(*j*) pojačavaju utisak percepcije stridensa. Snagu ovih spektralnih komponenti označićemo sa *in-band power* $P_{inBand}(t,j)$. Sa druge strane, spektralne komponente van ovog frekvencijskog opsega pojačavaju efekat maskiranja stridensa. Snagu ovih spektralnih komponenti označićemo sa *out-band power* $P_{outBand}(t,j)$. *In-band* i *out-band* snage se računaju prema

$$P_{inBand}(t,j)=\frac{1}{2\Delta+1}\sum_{l=j-\Delta}^{j+\Delta} z^{(A)}(t,l)\,, \tag{7.5}$$

$$P_{outBand}(t,j)=\frac{1}{M-2\Delta-1}\sum_{l\notin\{j-\Delta,j+\Delta\}} z^{(A)}(t,l)\,, \tag{7.6}$$

gde je Δ celobrojna konstanta kojom se definiše broj susednih kanala. Element $z_m^{(A)}(t,j)$ izlazne matrice sa izraženim spektralnim vrhovima računa se prema

$$z_m^{(A)}(t,j) = \alpha(t,j) z^{(A)}(t,j), \tag{7.7}$$

gde su težinski koeficijenti $\alpha(t,j)$ dati kao

$$\alpha(t,j) = \min\left(P_{inBand} / P_{outBand}, 1\right). \tag{7.8}$$

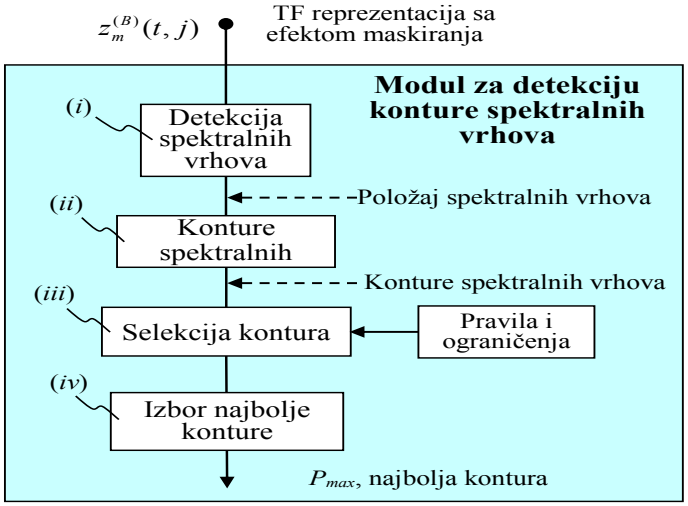
Izlaz $z_m^{(A)}(t,j)$ se koristi kod koraka 3 za izračunavanje auditorne slike - *strobed auditory image* (Patterson i Allerhand, 1995; Patterson, 2000) odabrane konture spektranih vrhova. Svrha grane (B) prikazane na slici 7.6 jeste da pruži vremensko-frekvencijsku predstavu govornog signala koja je pogodna za selekciju spektralnih vrhova. U ovu svrhu nije moguće koristiti $z_m^{(A)}(t,j)$ jer se sastoji samo od pozitivnih poluperioda, koje nisu pogodne za detekciju spektralnih vrhova. Vremensko frekvencijska predstava data u grani (B) mnogo je pogodnija za ovu svrhu. Faze procesiranja u grani (B) identične su onima u grani (A) osim faze (*ii*) u kojoj je HWR - polutalasni ispravljač zamenjen proračunom apsolutne vrednosti (intenzitetom) $f(z)=|z|$,

$$x^{(B)}(t,j) = |s(t,j)| = \sqrt{\mathrm{Re}\left(s(t,j)\right)^2 + \mathrm{Im}\left(s(t,j)\right)^2} \tag{7.9}$$

gde su Re(.) i Im(.) realni i imaginarni deo kompleksne vrednosti, a izlaz $x^{(B)}(t,j)$ je (glatka) anvelopa j-og kanalnog signala. Posle faza (*iii*)-(*v*) u grani (B), avelope kanalnih signala $z_m^{(B)}(t,j),\ j=1,\cdots,Nch$ ostaju glatke, pogodne za selekciju kontura spektralnih vrhova u koraku 2 blok dijagrama prikazanog na slici 7.5.

7.1.1.3 Izbor konture spektralnih vrhova

Blok dijagram modula za nalaženje kontura spektralnih vrhova prikazan je na slici 7.7.



Slika 7.7 Modul za izdvajanje optimalne trajektorije spektralnih vrhova sa stanovišta mogućeg prisustva stridensa.

Ovim modulom se analizira matrica $z_m^{(B)}(t,j)$ i na njoj nalaze sve konture spektralnih vrhova koje mogu biti percepirane kao stridens. Ovo se realizuje kroz četiri koraka. U prvom koraku (*i*) za svaki vremenski indeks *t*, nalaze se i notiraju svi spektralni vrhovi. U narednom koraku (*ii*) spektralni vrhovi se povezuju u konture koristeći sledeće pravilo: ako je (*t*,*j*) poslednja tačka neke od kontura do trenutka *t* na kanalu *j*, i ako u trenutku *t*+1 neka od tačaka (*t*+1,*j*-1), (*t*+1,*j*), (*t*+1,*j*-1) predstavlja lokalni maksimum, tada se kontura produžava i na tačku (t+1,$j_{\max}$), gde je $j_{\max} = \arg\max_l \left(z_m^{(B)}(t,l)\right), l \in (j-1, j, j+1)$. U suprotnom, kontura se prekida. Lokalni maksimumi u trenutku *t*+1 koji nisu obuhvaćeni nekom od kontura iz trenutka *t*, predstavljaju početke novih konture.

Sve konture ne mogu predstavljati stridens. Selekcija kontura koje mogu da budu percepirane kao stridens vrši se u modulu (*iii*). Te konture treba da ispune sledeća dva uslova (ograničenja):

(c_1) Da je njihovo trajanje veće od minimalnog vremenskog intervala T_{min}. Trajanje minimalnog vremenskog intervala je određeno eksperimentalno i on iznosi T_{min}=9ms. Konture kraćeg trajanja od T_{min} se ne percepiraju kao stridens.

(c_2) Pseudoperiodičan signal stridensa treba da ima stabilnu rezonantnu učestanost, koja je ili konstantna (primer slika 7.1a, slika 7.1b prva rezonantna linija) ili se linearno menja u vremenu (primer slika 7.1b druga rezonantna linija, slika 7.1c). U suprotnom, taj deo signala neće proizvesti utisak stridensa već frikcije. Ako trajektoriju

spektralnih vrhova aproksimiramo pravom linijom, mera nestabilnosti je srednje kvadratno odstupanje u odnosu na aproksimaciju. U prvoj aproksimaciji ovog modela smatraćemo da je učestanost stridensa konstantna na kratkom vremenskom intervalu trajanja T_{min}. Mera nestabilnosti stridensa je tada standardna devijacija na prozoru analize trajanja T_{min}. Da bi odgovarajući segmenat proizveo utisak stridensa, standardna devijacija na klizećem prozoru trajanja T_{min} treba da je manja od unapred definisanog praga odluke σ_λ.

Konture koje ispunjavaju uslove (c_1) i (c_2) su potencijalni izvori stridensa. Niz tačaka koje odgovaraju konturi p_i obeležićemo sa $S_i^{(B)}(t)$,

$$S_i^{(B)}(t) = z_m^{(B)}(t, p_i(t)), \quad t = t_1,...,t_{end}\,, \tag{7.10}$$

gde je t_1 prva a t_{end} poslednji element konture p_i. Sa $Str_i^{(B)}(t)$ definiše se lokalna mera snage konture p_i u diskretnom vremenskom trenutku t sa

$$Str_i^{(B)}(t) = \begin{cases} \sum_{j=t}^{t+L-1} S_i^{(B)}(j), & za \;\; std(p_i(t),..., p_i(t+L-1)) < \sigma_\lambda \\ 0 & inace \end{cases}, \tag{7.11}$$

gde *std*(.) predstavlja standardnu devijaciju, a L dužinu MA filtra čiji je početak postavljen u T_{min}. Prag varijanse σ_λ treba biti određen eksperimentalno, što je i učinjeno i postavljen je na 1.

Lokalna mera snage $Str_i^{(B)}(t)$ suma je elemenata $S_i^{(B)}(j),\; j = t, t+L-1$ ukoliko važi uslov c_2. U suprotnom, ona je jednaka nuli jer ovaj deo konture sa elementima $(t, p_i(t)),...,(t, p_i(t+L-1))$ ne može biti percipiran kao stridens već kao šum. Meru snage konture p_i obeležimo sa

$$Str_i^{\max} = \max\left(Str_i^{(B)}(t), t = t_1,...,t_{end} - L + 1)\right). \tag{7.12}$$

Kontura sa maksimalnom merom snage obeležena je sa $pi_{\max}$ a izabrana je kao ona sa najvećom šansom da bude percepirana kao stridens. Indeks ove konture dat je sa

$$i_{\max} = \arg\max_i \left(Str_i^{\max}\right) \tag{7.13}$$

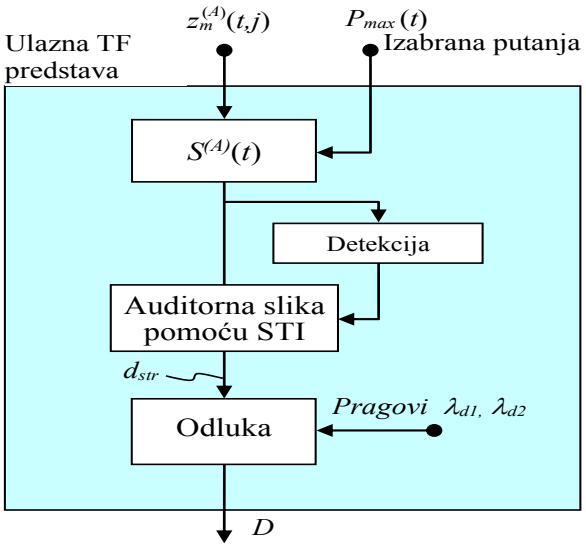
A kontura p_{max}, kao

$$p_{max} = pi_{\max}\,, \tag{7.14}$$

U narednom modulu za detekciju stridensa na bazi STI testira se da li odabrana kontura p_{max} zaista proizvodi stridens na osnovu posebno definisane mere subjektivnog osećaja stridensa.

7.1.1.4 Detekcija stridensa na bazi STI (strobed temporal integration)

Blok dijagram koraka obrade u kome se vrši detekcija stridensa prikazan je na slici 7.8. Detekcija se bazira na dvema ulaznim veličinama. Prva je vremensko - frekvencijska matrica $z_m^{(A)}(t, j)$ dobijena *Patterson*-ovim auditornim modelom. Druga ulazna veličina je kontura spektralnih vrhova p_{max} koja ima najveću šansu da bude percepirana kao stridens.



Slika 7.8 Modul za detekciju prisustva stridensa.

Za razliku od *Paterson*-ovog postupka gde se elementi auditorne slike formiraju za svaki od kanalnih signala nezavisno (Patterson and Holdsworth, 1996), za potrebe detekcije stridensa se formira auditorna slika duž odabrane konture p_{max}. Formiranje auditorne slike se sastoji iz više koraka. U prvom koraku se formira niz tačaka u TF predstavi signala $z_m^{(A)}(t, j)$ duž odabrane konture $S^{(A)}(t)$,

$$S^{(A)}(t) = z_m^{(A)}(t, p_{\max}(t)), \quad t = t_1, \ldots, t_{end} \tag{7.15}$$

Primer tipičnog niza $S^{(A)}(t)$ je prikazan na slici 7.9a. Za razliku od niza $S_i^{(B)}(t)$ koji je gladak jer je formiran na osnovu anvelopa kanalnih signala $z_m^{(B)}(t,j)$ (vidi jednačinu (7.10)), niz tačaka $S^{(A)}(t)$ je povorka pozitivnih poluperioda koje pored informacije o snazi nose i informaciju o faznom stavu kanalnih signala.

Slično postupku formiranja auditorne slike metodom stroboskopske vremenske integracije (STI), (Patterson i Holdsworth, 1996; Patterson, 2000) u narednom koraku se vrši pozicioniranje okidajućeg impulsa *T(k), k*=1,...,N_{tr} na lokalnim maksimumima pozitivnih poluperioda. N_{tr} je ukupan broj triger impulsa koji se formiraju na odabranoj konturi spektralnih vrhova p_{max}. Lokalni maksimumi se detektuju na osnovu premašenja adaptivnog praga odluke $\alpha(t)$, na slici 7.9a označenog tankom linijom. Prag $\alpha(t)$ eksponencijalno opada sa vremenom, a nakon detekcije triger impulsa, setuje se na vrednost lokalnog maksimuma signala $S^{(A)}(t)$, (Patterson i Holdsworth, 1996; Patterson, 2000). Pozicije lokalnih maksimuma su na slici 7.9a označeni rombovima. U trenutku aktiviranja triger impulsa (koji se nalaze na pozicijama lokalnih maksimumima), aktivira se akumuliranje signala u auditornom baferu postupkom opisanim pseudokodom :

for $k=1,\ldots,N_{tr}$

$A_{buf}(N-T(k)+1:N)=A_{buf}(N-T(k)+1:N)+S^{(A)}(1:T(k))$

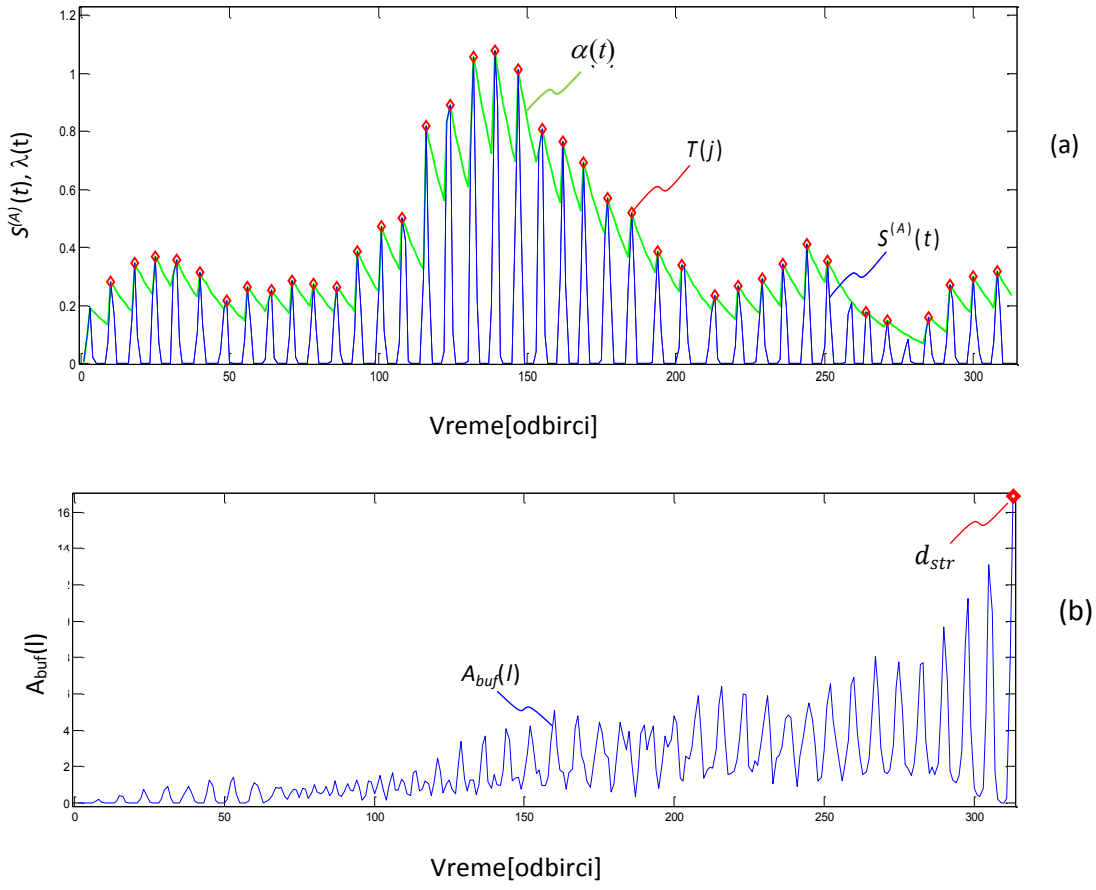
end

Vektor $A_{buf}(l)$, l=1,..,N je auditorni bafer koji sadrži ukupno N tačaka. Sadržaj auditornog bafera nakon primene navedenog STI postupka prikazan je na slici 7.9b. Za naš algoritam od važnosti je samo poslednji elemenat auditornog bafera $A_{buf}(N)$ koji koristimo kao meru prisustva stridensa, $d_{str} = A_{buf}(N)$.

Odluka o prisustvu stridensa se donosi poredeći meru stridensa d_{str} sa dva praga odluke λ_{d1}, λ_{d2}, $0<\lambda_{d1}<\lambda_{d2}$ prema pravilu

$$D = \begin{cases} D_0 & bez\ stridensa, \quad for\ d_{str} < \lambda_{d1} \\ D_1 & sa\ stridensom, \quad for\ \lambda_{d2} > d_{str} \\ D_2 & bez\ odluke, \quad for\ \lambda_{d1} \le d_{str} \le \lambda_{d2} \end{cases} \tag{7.16}$$

Odluka D_0 ima značenje da u izgovoru subjekta nije prisutan stridens. Odluka D_1 znači da je u izgovoru subjekta prisutan stridens. Odluka D_2 znači da se funkcija odluke d_{str} nalazi u oblasti u kojoj ne možemo sa sigurnošću da tvrdimo ni da ima ni da nema stridensa.



Slika 7.9 Stroboskopska vremenska integracija
(a)Debela linija - niz $S^{(A)}(t)$, tanka linija - Vremenski promenjiv prag odluke $\alpha(t)$, rombovi – pozicije (okidajućeg) triger impulsa; (b)Sadržaj auditornog bafera $A_{buf}(l)$. Rombom je označen poslednji odbirak auditornog bafera koji predstavlja izračunatu vrednost stridensa d_{str}.

7.1.1.5 Rezultati

Za testiranje predloženog modela korišćena je baza snimaka dece školskog uzrasta starosti 10 - 11 godina. Deca posmatranog uzrasta treba da imaju usvojene sve izgovorne glasove srpskog jezika i da ih pravilno izgovaraju u svim fonološkim pozicijama. Međutim, u ovoj populaciji ima dosta dece sa nepravilnim izgovorom jednog ili više glasova među kojima su najzastupljeniji glasovi iz grupe frikativa. Analizirajući akustičke karakteristike stridensa u posmatranom uzorku uočeno je da se on javlja na frekvencijama iznad 1000 Hz. Kako zvučnost fonema u ovom frekvencijskom opsegu ne utiče na performanse algoritma, primenjeni postupak se može jednako primeniti i na zvučne i na bezvučne foneme. Stoga je algoritam testiran na frikativu /š/ koji se javlja kao najčešće nepravilno izgovoreni glas.

Da bi se ocenila uspešnost algoritma za detekciju stridensa iz baze snimaka izdvojene su tri grupe subjekata:

(a) 16 subjekata kod kojih je jedino patološko obeležje izgovora /š/ bilo stridens različitog stepena od blagog do izrazito naglašenog,

(b) 16 subjekata kod kojih je izgovor /š/ bio korektan (bez odstupanja).

(c) 16 subjekata sa izgovorom /š/ bez stridensa ali sa nekim drugim oblicima patološkog izgovora.

Tako je dobijena test baza od ukupno 48 subjekata koji su korišćeni za testiranje algoritma.

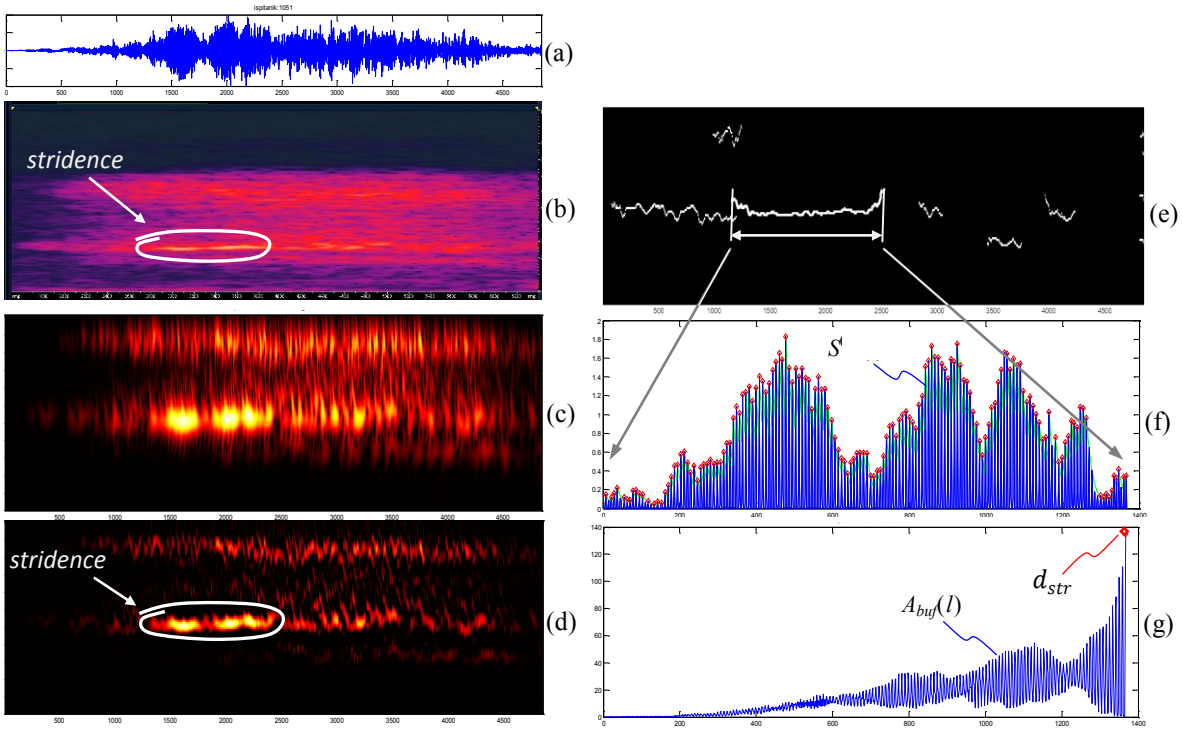
Testiranje algoritma je izvršeno na reči /šuma/, pri čemu je analizirano prisustvo stridensa na inicijalnoj poziciji frikativa /š/. U eksperimentima u kojima eksperti ocenjuju kvalitet govora ili glasa često se opravdano postavlja pitanje njihove stručnosti i sposobnosti da uoče određene akustičke detalje. U literaturi postoje suprotstavljena mišljenja o pouzdanosti rezultata dobijenih na osnovu jednog ili više eksperata (Kent, 1996). U ovom eksperimentu za odluku o prisustvu stridensa korišćeni su rezultati ocenjivanja od strane jednog iskusnog eksperta. Za ovakav pristup smo se opredelili imajući u vidu da iskustvo može imati značajnu ulogu u sposobnosti ocenjivača da uoči suptilne razlike u kvalitetu izgovora jedne foneme ali i da eksperti mogu imati veoma različita iskustva i perceptivne sposobnosti. Opredelili smo se za jednog eksperta iz

Instituta za eksperimentalnu fonetiku i patologiju govora u Beogradu, iz sledećih razloga:

1. Odabrani ekspert se preko 20 godina bavi govorno jezičkom terapijom,
2. Odabrani ekspert svakodnevno procenjuje artikulaciju pacijenata različitog uzrasta i patologije (godišnje oceni između 600-800 izgovora različitih pacijenata)
3. Odabrani ekspert pokazuje najveću pouzdanost ocenjivanja u različitim psihoakustičkim testovima provere kvaliteta artikulacije.

Baza govornih stimulusa je snimljena sa učestanošću odabiranja 44100 Hz. U predobradi je vršena decimacija signala sa 2 tako da je konačna učestanost odabiranja govornog signala bila 22050 Hz. Gammatone banka filtara je sadržala 111 kanala sa centralnim učestanostima raspoređenim po ERB skali u opsegu od 900 Hz do 7940 Hz. Opseg 900 – 7940 Hz je odabran iz razloga što se u ovom opsegu najčešće javlja stridens.

Na slici 7.10 su prikazani međurezultati obrade tipičnog izgovora frikativa /š/ sa stridensom. Na slikama 7.10a, 7.10b i 7.10c prikazani su redom vremenski dijagram signala, FFT spektrogram i TF prikaz preko banke filtara. Na slici 7.10d je prikazana TF reprezentacija signala $z_m^{(B)}(t,j)$ dobijena primenom postupka isticanja spektralnih vrhova (jednačine (7.3), (7.4)) i simulacije efekta maskiranja (jednačine (7.5), (7.6), (7.7), (7.8)). Na slici 7.10e su tankim belim linijama prikazane konture koje zadovoljavaju uslove (c_1) i (c_2). Odabrana kontura sa najvećom merom stridensa je prikazana debelom belom linijom. Na slici 7.10f su punom linijom prikazani elementi niza $S^{(A)}(t), t = t_1,...,t_{end}$ koji se odnose na odabranu konturu (p_{max}) prikazanu debelom linijom na slici 7.10e. Postupak *strobed temporal integration* (STI) (Patterson, 2000; Patterson i Allerhand, 1995) opisan pseudokodom u tabeli T1, primenjuje se na vektor $S^{(A)}(t), t = t_1,...,t_{end}$. Položaji okidajućeg impulsa su označeni rombovima. Na slici 7.10g je prikazan sadržaj auditornog bafera $A_{buf}(l)$ odabrane konture p_{max} gde poslednji element auditornog bafera označen rombom predstavlja brojnu vrednost mere stridensa d_{str}.



Slika 7.10 Frikativ /š/ sa stridensom.
(a) Vremenski dijagram fonema /š/ u reči /šuma/, (b) FFT spektrogram, (c) kohleagram, (d) kohleagram sa istaknutim spektalnim vrhovima, (e) tanka linija- konture spektralnih vrhova, debela linija – odabrana kontura spektralnih vrhova ,(f) $S^{(A)}(t)$ duž odabrane konture p_{max}, (jednačina 7.15), gde je pozicija trigera označena rombom, (g) sadržaj auditornog bafera $A_{buf}(l)$. Poslednji element auditornog bafera označen rombom predstavlja meru stridensa d_{str}

Za svakog ispitanika iz (a), (b) i (c) sračunata je mera distance d_{str} koristeći pseudo kod dat ranije.

Raspodela mere stridensa d_{str} za pacijente iz grupa (b) i (c), (oni bez stridensa), sračunata je kao broj ispitanika bez stridensa sa merom stridensa većom od usvojenog praga α kao

$$F_{ws}(d_{str}) = \sum_{i=1}^{N_{ws}} \left(\begin{cases} 1, & za \;\; d_{str} < d_{str}(i) \\ 0, & ostalo \end{cases} \right) \tag{7.17}$$

gde je $F_{ws}(d_{str})$ raspodela mere stridensa $d_{str}(i)$ subjekata bez stridensa, i je redni broj subjekta, a N_{ws} je ukupan broj subjekata bez stridensa. Raspodela mere d_{str} za subjekte sa stridensom, grupa (a), izračunava se kao broj subjekata sa stridensom sa merom manjom od argumenta d_{str} primenjujući formulu

$$F_{str}(d_{str}) = \sum_{i=1}^{N_s} \left(\begin{cases} 1, & za \;\; d_{str}(i) < d_{str} \\ 0, & ostalo \end{cases} \right) \tag{7.18}$$

gde je N_s je ukupan broj subjekata sa stridensom, $F_{str}(d_{str})$ je raspodela mere stridensa subjekata sa stridensom, a $d_{str}(i)$ je mera stridensa subjekta i, i=1,...,N_s.

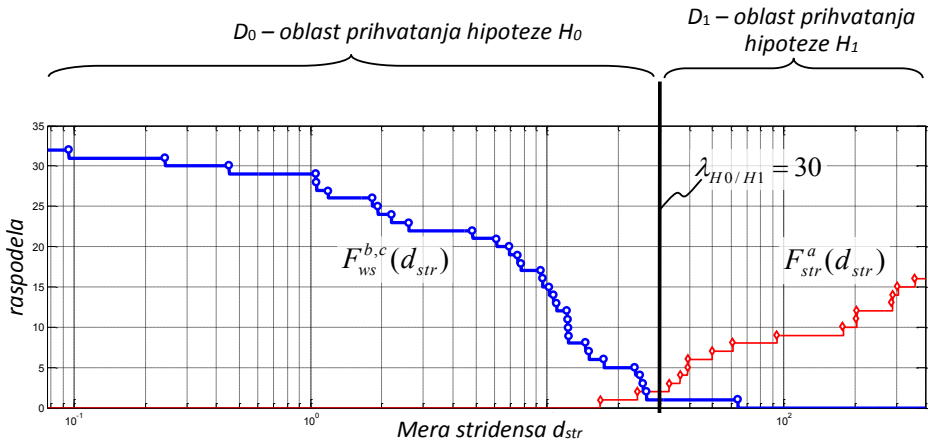
Testiranje pouzdanosti detekcije je ocenjivano kroz tri test slučaja.

U prvom test slučaju klasu subjekata bez stridensa činili su subjekti iz grupe (b) – subjekti bez ikakve govorne patologije i subjekti iz grupe (c) – subjekti bez stridensa ali sa nekim drugim oblikom patološkog izgvora, ukupno 32 subjekta. Drugu klasu subjekata sa stridensom činili su subjekti iz grupe (a), ukupno 16 subjekata. Na slici 7.11 tankom linijom je prikazana raspodela mere stridensa $F_{str}(d_{str})$ subjekata sa stridensom iz grupe (a). Na istoj slici je debelom linijom prikazana raspodela mere stridensa $F_{ws}(d_{str})$ subjekata bez stridensa iz grupa (b) i (c).

Kao klasifikator prisustva stridensa primenjen je kNN klasifikator sa 3 najbliža suseda. Za taj tip klasifikatora smo se opredelili zbog uspešnosti njegove primene u slučajevima kada se ne poznaju raspodele pod hipotezama H_0 i H_1 i zbog toga što za njegovu primenu nije potrebna prethodna obuka. S obzirom da se kod ovog klasifikatora odluka donosi na osnovu svih uzoraka isključujući testirani uzorak, može se reči da je ovaj postupak sličan postupku *leave-one-out cross-validation* (LOOCV) (Kohavi, 1995; Cawley i Talbot, 2003). Primenom ovog postupka klasifikacije broj lažnih detekcija stridensa je iznosio $\varepsilon_{H1|H0} = 1/32$ (jedna lažna detekcija na ukupno 32 subjekata bez stridensa). Hipoteza H_1 znači prisutan stridens nasuprot alternativnoj hipotezi H_0 da stridens nije prisutan. Broj propuštenih detekcija stridensa je iznosio $\varepsilon_{H0|H1} = 2/16$ (dve propuštene detekcije od ukupno 16 subjekata sa stridensom).

Primena kNN za klasifikaciju u ovom testnom slučaju definisane su 2 oblasti: D_0 (za $d_{str} < \lambda_{H0/H1}$), oblast prihvatanja hipoteze H_0, i D_1 (za $d_{str} \geq \lambda_{H0/H1}$) oblast prihvatanja hitpoteze H_1. Međutim postavljanje ovako oštre granice između oblasti prihvatanja hipoteza H_0 i H_1 u slučaju detekcije stridensa nije u skladu sa procesom kategorijalne percepcije ili kategorizacije govora (Holt i Lotto, 2010). Donošenje odluke o pripadnosti jednoj ili drugoj klasi možemo posmatrati kroz propabilistički pristup (Nosofsky, 1992) u kome su verovatnoće prepoznavanja u zonama jasnog razdvajanja klasa bliske 1. Između ove dve zone postoji prelazna oblast ili oblast nesigurnosti u

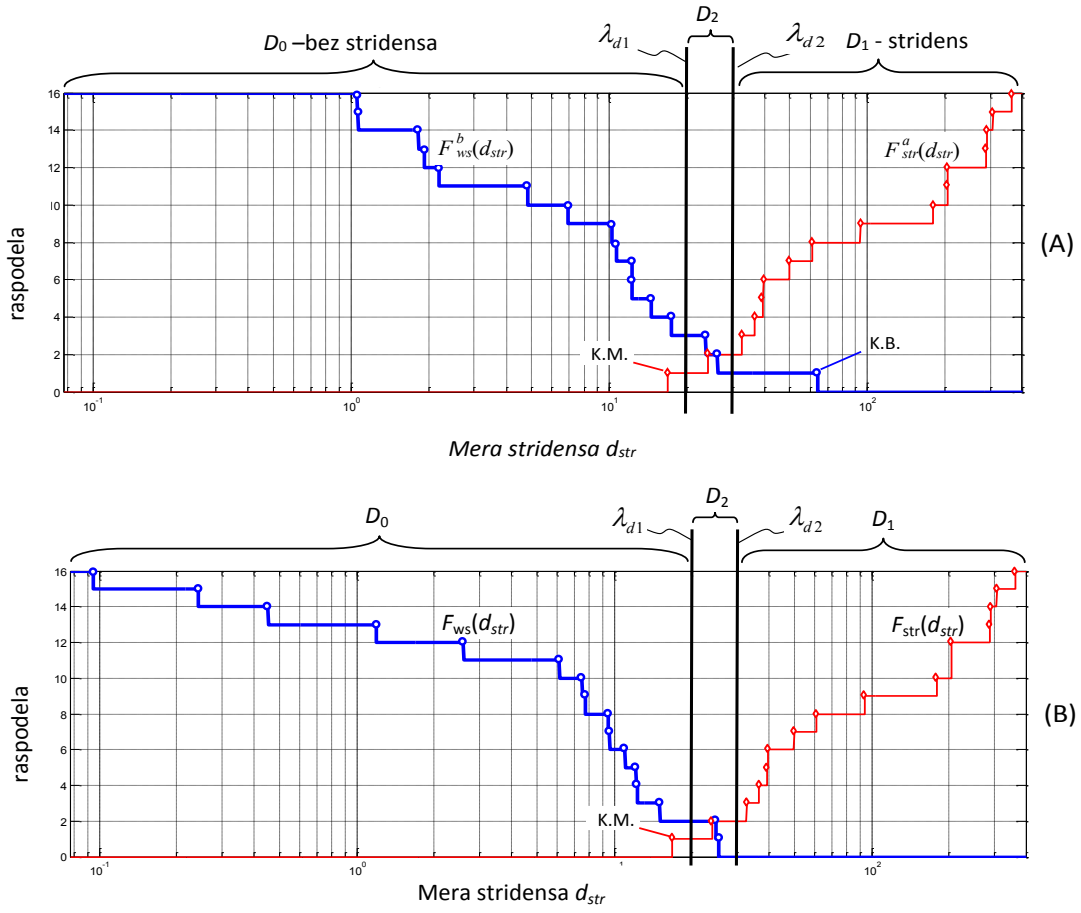
kojoj je verovatnoća klasifikacije slučajna. Širina ove oblasti zavisi od iskustva (Livingston i sar., 1998). Imajući ovo u vidu algoritam je modifikovan tako da su definisana dva praga odluke λ_{d1} i λ_{d2}, $\lambda_{d1} < \lambda_{d2}$ kojima se definišu tri oblasti odluke. Skup vrednosti mere stridensa d_{str} ,$d_{str}>\lambda_{d2}$ definiše oblast usvajanja odluke da je stridens prisutan koju ćemo označiti sa D1. Skup vrednosti d_{str}, $d_{str}<\lambda_{d1}$ definiše oblast usvajanja odluke da stridens nije prisutan. Ovu oblast ćemo označiti sa D0. I konačno, skup vrednosti d_{str}, $\lambda_{d1} \le d_{str} \le \lambda_{d2}$ definišu oblast D2, oblast u kojoj ne možemo doneti pouzdanu odluku o stridensu. Za prag odluke λ_{d2} , usvojili smo vrednost praga odluke iz test slučaja 1, odnosno $\lambda_{d2} = \lambda_{H0/H1} = 30$. Vrednost praga odluke λ_{d1} smo birali tako da izjednačimo broj propuštenih detekcija sa brojem lažnih detekcija. Na osnovu tog kriterijuma je usvojeno λ_{d1}=20. Ovako definisane pragove smo koristili u test slučajevima 2 i 3.



Slika 7.11 Testni slučaj 1. Distribucija mere stridensa: debela linija - $F_{ws}^{b,c}(d_{str})$ za subjekte sa korektnim izgovorom frikativa /š/ bez stridensa iz grupa (b) i (c), tanka linija - $F_{str}^{a}(d_{str})$ za subjekte sa stridensom – grupa (a). Prag odluke $\lambda_{H0/H1}$=30 dobijen je primenom kNN klasifikatora sa 3 najbliža suseda.

U test slučaju 2 su korišćeni snimci subjekata iz grupe (a) i iz grupe (b), i već određeni pragovi λ_{d1} i λ_{d2}. Raspodele mere stridensa pod hipotezom H0, grupa (b) i hipotezom H1, grupa (a) prikazane su redom debelom i tankom linijom na slici 7.12(A). Iz grupe (b), subjekti bez stridensa, lažno je detektovan jedan subjekat kao subjekat sa stridensom $\varepsilon_{H1|H0} = 1/16$. Iz grupe (a), subjekti sa stridensom, pogrešno je klasifikovan jedan subjekat kao subjekat bez stridensa $\varepsilon_{H0|H1} = 1/16$. U oblasti D2 u kojoj nije

moguće doneti pouzdanu odluku bilo je 3 od ukupno 32 subjekata iz grupa (a) i (b) $\varepsilon_{D2} = 3/32$. U test slučaju 3 subjekti bez stridensa su imali neki drugi oblik patološkog izgovora, grupa (c). U ovom test slučaju je korišćena ista grupa subjekata sa stridensom kao i u prethodnom, grupa (a). Na slici 7.12(B) debelom i tankom linijom su prikazane raspodele mere stridensa redom za grupe (c) i (a). Za iste vrednosti pragova λ_{d1} i λ_{d2} ni jedan subjekat iz grupe (c) nije pogrešno klasifikovan. Iz grupe (a) propuštena je detekcija jednog od ukupno 16 subjekata sa stridensom $\varepsilon_{H0|H1} = 1/16$. U oblasti D2 u kojoj nije moguće doneti pouzdanu odluku bilo je ukupno 3 subjekta iz grupa (a) i (c), $\varepsilon_{D2} = 3/32$.



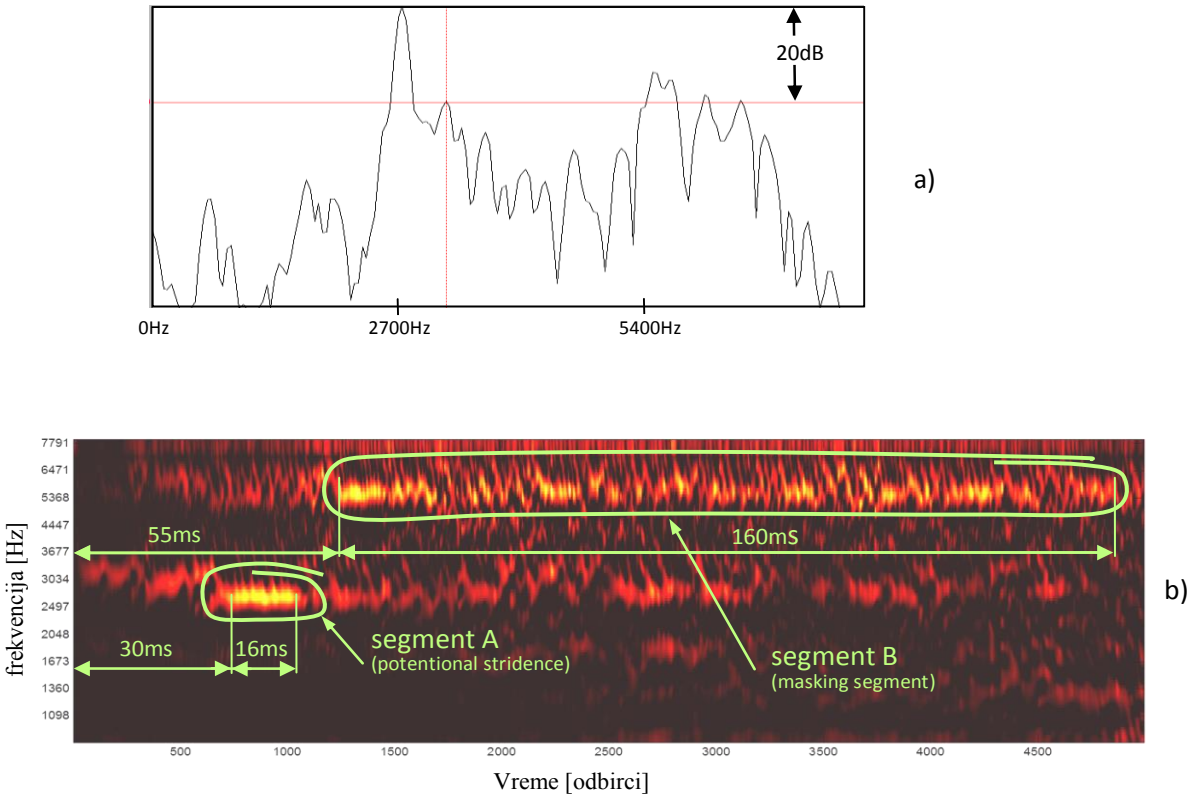
Slika 7.12 Distribucija mere stridensa, test slučaj 2 i 3.
(A) Testni slučaj 2. Distribucija mere stridensa: debela linija - $F^b_{ws}(d_{str})$ za subjekte sa korektnim izgovorom frikativa /š/ bez stridensa, tanka linija - $F^a_{str}(d_{str})$ za subjekte sa stridensom. (B) Testni slučaj 3. Distribucija mere stridensa: debela linija - $F_{ws}(d_{str})$ za subjekte sa patološkim izgovorom frikativa /š/ koje nije stridens, tanko - $F_{str}(d_{str})$ za subjekte sa stridensom.

7.1.1.6 Diskusija i zaključci

Analiza rezultata dobijenih predloženim modelom pokazuje da se javljaju dva slučaja nepravilne klasifikacije eksperimentalnih rezultata. Prvi je propuštena detekcija a drugi lažna detekcija stridensa.

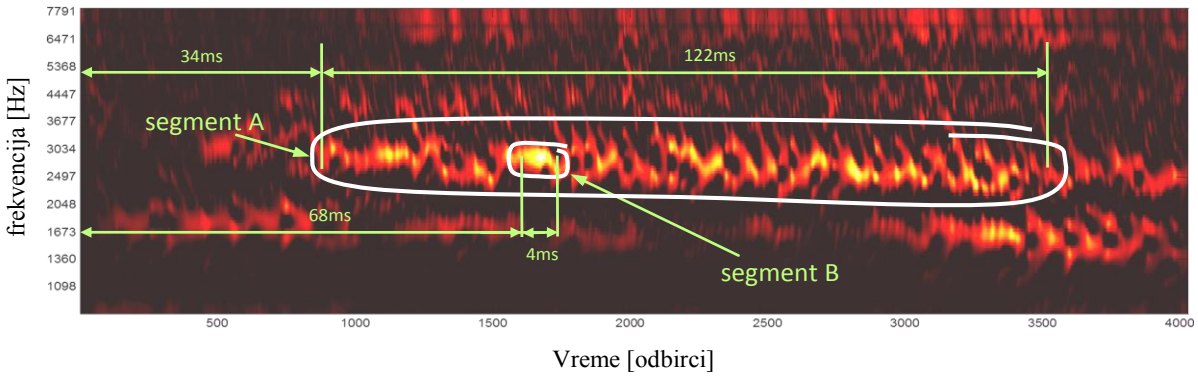
Primer lažne detekcije stridensa je subjekat označen sa K.B. (Slika 7.12(A)). Ekspert ocenjivač je ocenio da subjekat K.B. nema stridens, dok je algoritam detektovao stridens. Na slici 7.13b je prikazan kohleagram sa istaknutim spektralnim vrhovima frikativa /š/ i označenim karakterističnim segmentima A i B. Uočava se da su segmenti dislocirani i u vremenom i u frekvencijskom domenu. Na segmentu A (30-46 ms) jasno se uočava kontura spektralnih vrhova u okolini učestanosti 2700 Hz. Na slici 7.13a je prikazan usrednjen spektar segmenta A gde se jasno vidi da je spektralni vrh za oko 20 dB iznad okolnih spektralnih komponenti. To upućuje na moguće prisustvo stridensa. Izračunata mera stridensa na segmentu A, d_{str} =64.3, veća je od praga odluke $\lambda_{d2} = 30$, zbog čega se donosi odluka da je stridens prisutan. Uprkos tome stridens se ne čuje a razlog tome su spektralne komponente na segmenat B (55-215 ms) koje maskiraju stridens. Ove spektralne komponente nemaju obeležje stridensa već pojačane frikcije. Da bismo potvrdili pretpostavku da segmenat B maskira segmenat A, na frikativ /š/ smo primenili filter nepropusnik učestanosti 4000-7000 Hz sa slabljenjem od 40 dB u nepropusnom opsegu. Tako obrađen signal ponovo smo dali na ocenu ekspertu ocenjivaču. U ovom drugom slučaju ekspert je konstatovao prisustvo stridensa, čime je potvrđena naša pretpostavka o maskiranju stridensa od strane segmenta B.

U ovom test slučaju javlja se vremensko maskiranje, odnosno maskirajući i maskirani signal se ne preklapaju u vremenu. Ovaj tip maskiranja je maskiranje u nazad, jer se maskirani signal završava 9 ms pre nego što se javlja maskirajući signal. (videti sliku 7.13b). Sa dijagrama prikazanog na slici 10.16 u referenci Moore (1997) strana 198, prag maskiranja za 9 ms je negde oko 12 dB što efektivno smanjuje odnos signal šum na oko 8 dB. Prema potrebnim kriterijumima za odnos signal/šum mora biti veći od 20 dB (Fuchs, 2007) da bi stridens bio percipiran.



Slika 7.13 Primer subjekta bez stridensa koji se pogrešno klasifikuje kao stridens: a) usrednjeni spektar segmenta A, b) kohleagram.

Druga karakteristična greška je propuštena detekcija stridensa kod subjekta G.P. koji je od strane ocenjivača ocenjen da ima stridens. Na slici 7.14. prikazan je kohleagram sa markiranim spektralnim segmentom A sa osnovnom karakteristikom pojačane frikcije. U ovoj frikciji nalazi se niz izolovanih spektralnih impulsa koji imaju karakteristike stridensa (vidi sliku 7.2). Međutim, oni su veoma kratkog trajanja (najintenzivniji segment B je trajanja 4 ms) i zbog toga ni jedan od ovih vrhova algoritam nije mogao detektovati kao stridens. Sa druge strane, ekspert je perceptivno detektovao stridens. Ova pojava se može objasniti pomoću perceptivne restauracije nedostajućeg zvuka (Warren, 2008.). Naime, između stridentnih impulsa nalazi se pojačana frikcija koja omogućava perceptivnu restauraciju stridensa. Ova iluzija kontinuiteta ilustruje konstruktivnu prirodu percepcije koja ekspertima omogućava da percepiraju stridens. Ovaj psihoakustički efekat nije uključen u ovoj verziji algoritma i zbog toga je došlo do propuštene detekcije stridensa.



Slika 7.14 Primer subjekta sa stridensom koji je pogrešno klasifikovan kao subjekat bez stridensa.

Prethodna analiza govori u prilog tome da određeni psihoakustički efekti u mnogome utiču na percepciju akustičkih obeležja patološkog glasa. Kako je efekat maskiranja ugrađen u opisani algoritam, postavlja se pitanje na koji način i u kojoj meri on utiče na detekciju stridensa. Pored toga, potrebno je i utvrditi kako upotreba auditornih modela u algoritamskom pristupu detekciji stridensa utiče na detekciju istog.

7.1.2 Detekcija stridensa na osnovu FFT spektra i Burgovog algoritma

U ovom poglavlju opisan je algoritam za detekciju stridensa, koji predstavlja preteču algoritma prethodno opisanog u poglavlju 7.1.1. Kod ovog algoritma za modelovanje govornog signala koristi se FFT spektar, a za određivanje prisustva stridensa Burgov algoritam, tako da on, za razliku od prethodno opisanog algoritma, ne obuhvata psihoakustičke efekte kao ni modelovanje govornog signala pomoću auditornog modela. Poređenjem ova dva algoritma može se utvrditi u kom obimu korišćenje auditornog modela i psihoakustičkih efekata poboljšava automatsku detekciju stridensa u izgovoru.

U ovom pristupu detekcije stridensa koristi se vremensko frekvencijska predstava govornog signala dobijena FFT-om (Jovičić i sar., 2008). Algoritam postupka je prikazan na slici 7.16 i prilagođen je detekciji stridensa kod frikativa */š/,/ž/,/ć/ i /dž/*. Ulazni signal se najpre filtrira filtrom propusnikom opsega učestanosti BP(f_1, f_2) sa granicama koje zavise od foneme koja se analizira. Granice filtara date su u tabeli 7.1.

Tabela 7.1 Granice upotrebljenih filtara

Fonema	f_1[Hz]	f_2[Hz]
š, ž	2000	4000
ć, dž	3000	8000

Tako filtriran signal deli se na preklapajuće segmente veličine 512 odbiraka (govorni signal je uzet sa frekvencijom odabiranja od 22050 Hz), $X(L)$ gde je sa L označen redni broj segmenta. Dalje procesiranje signala vrši se iz dva koraka. U prvom koraku se vrši određivanje potencijalnih pozicija segmenata sa stridensom dok se u drugom koraku donosi odluka o postojanju stidensa (slika 7.15).

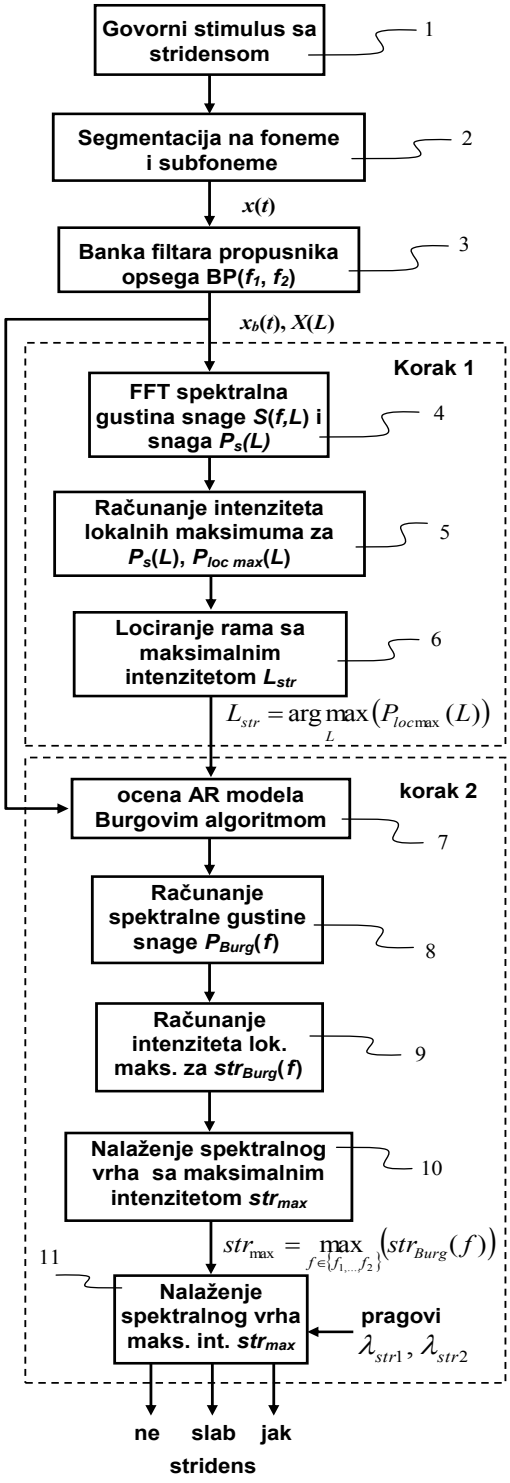
U prvom koraku najpre se vrši FFT analiza signala u svakom od segmenata i na taj način se dobija spektralna gustina snage $S(f,L)$ i snaga $P_s(L)$. Lokalni maksimumi sekvence $P_s(L)$ predstavljaju potencijalne pozicije stridensa pa je cilj pronaći sve lokalne maksimume $P_{loc\ max}(L)$, izračunati snagu i izdvojiti segment $L_{\max}$ koji predstavlja indeks segmenta sa najvećom vrednosti snage

$$L_{str} = \arg\max_{L}\left(P_{loc\max}(L)\right) \tag{7.19}$$

Kada se locira segment sa najvećom vrednošću snage, potrebno je proceniti da li se zaista radi o stridensu. Taj zadatak obavlja se u sledećem koraku procesiranja. U frekvencijskom domenu stridens se manifestuje kao intenzivna uskopojasna linija u spektru. Ovaj fenomen određuje se pomoću auto regresivnog (AR) modela procenom pomoću Burgovog algoritma (Burg, 1975; Burg, 1967). AR model šestog reda procenjuje se pomoću Burgovog algoritma na segmentu $X(L_{max})$ i izračunava se spektar snage $P_{Burg}(f)$, nakon čega se određuje snaga lokalnog spektralnog maksimuma $str_{Burg}(f)$ pomoću klizajućeg prozora. AR model šestog reda ima tri spektralna maksimuma, a maksimum vrednosti $str_{Burg}(f)$ dobija se kao $str_{\max} = \max_{f\in\{f_1,\ldots,f_2\}}\left(str_{Burg}(f)\right)$. Konačno, uvodi se funkcija granice u formi

$$str_D = \frac{1}{1+e^{-(str_{\max}-\lambda_{str})}} \tag{7.20}$$

gde je parametar λ_{str} eksperimentalno utvrđena vrednost praga. Ova veličina može uzeti dve vrednosti λ_{str1} i λ_{str2} i na taj način obezbediti tri izlazne vrednosti algoritma i to: odsustvo stridensa, slab stridens i izražen stridens.



Slika 7.15 Algoritam detekcije stridensa pomoću FFT spektra i Burgoviog algoritma

Algoritam je testriran na izgovoru 30-oro dece koji su uzeti iz govorne baze opisane u poglavlju 4.5, od kojih je 15 sa normalnim izgovorom, a preostalih 15 sa patološkim izgovorom u formi stridensa. U patološkoj grupi bilo je 8 izgovora glasa /š/, 4 izgovora glasa /ž/ i po jedan izgovor glasova /ć/ i /dž/. U svim slučajevima stridens je bio izražen. Rezultati su pokazali potpuno poklapanje algoritamske ocene sa ocenom datom od strane eksperta (logopeda).

7.1.3 Poređenje dva pristupa

Da bi se odredio uticaj (primenjenih) inkorporiranih psihoakustičkih efekata na detekciju stridensa uporedili smo rezultate dobijene algoritmom predloženim u poglavlju 7.1.1, u daljem tekstu označenog sa ALG_2014, i algoritmom opisanim u poglavlju 7.1.2 koji ne koristi auditorni model, u daljem tekstu ALG_2008. Skup ispitanika i testni slučajevi opisani su u poglavlju 7.1.1.6. Distribucija ispitanika bez stridensa za uzorke iz grupa (b) i (c) je računata sa

$$F_{ws}^{2008,(b|c)}(d^{2008}) = \sum_{i=1}^{N_{ws}} \left(\begin{cases} 1, & za \;\; d^{2008} < d^{2008}(i) \\ 0, & inace \end{cases} \right) \tag{7.21}$$

gde je $F_{ws}^{2008,(b|c)}(d^{2008})$ raspodela funkcije odluke $d^{2008}(i)$ dobijene algoritmom ALG_2008 za subjekte bez stridensa, i je redni broj subjekta, a N_{ws} je broj subjekata bez stridensa. Distribucija ispitanika sa stridensom za uzorke iz grupe (a) je računata sa

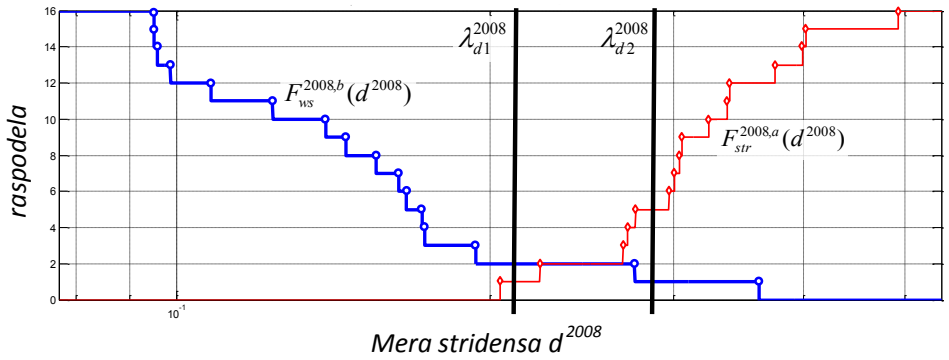
$$F_{str}^{2008,a}(d^{2008}) = \sum_{i=1}^{N_S} \left(\begin{cases} 1, & za \;\; d^{2008}(i) < d^{2008} \\ 0, & inace \end{cases} \right) \tag{7.22}$$

gde je $F_{str}^{2008,a}(d^{2008})$ raspodela funkcije odluke $d^{2008}(i)$ subjekata sa stridensom, a N_S je broj subjekata sa stridensom. Pragovi odluke $\lambda_{d1}^{2008} = 0.21$ i $\lambda_{d2}^{2008} = 0.28$ su određeni primenom kNN klasifikatora u test slučaju 1 na isti način kao što je to učinjeno u slučaju algoritma ALG_2014.

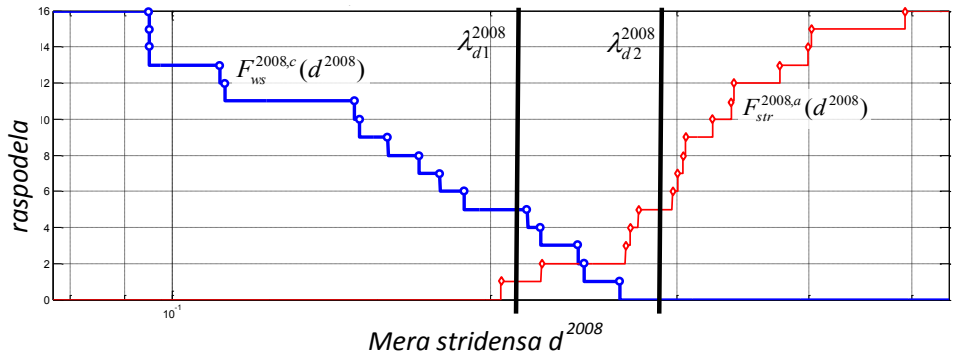
Na slici 7.17A prikazane su distribucije funkcije odluke algoritma ALG_2008 koje se odnose na test slučaj 2. Tankom crtom je prikazana raspodela mere stridensa

$F_{str}^{2008,a}(d^{2008})$ subjekata sa stridensom iz grupe (a), a debelom crtom raspodela mere stridensa $F_{ws}^{2008,b}(d^{2008})$ subjekata bez stridensa iz grupe (b).

U test slučaju 3 korišteni su subjekti iz grupa (a) i (c). Odgovarajuće raspodele $F_{str}^{2008,a}(d^{2008})$ i $F_{ws}^{2008,c}(d^{2008})$ prikazane su na slici 7.16B. U Tabeli 7.2 prikazani su rezultati detekcije stridensa dobijeni algoritmima ALG_2008 i ALG_2014.



(A)



(B)

Slika 7.16 Rezultati detekcije stridensa za algoritam ALG_2008 (A) Distribucija funkcije odluke d^{2008} algoritma ALG_2008 za test slučaj 2. (B) Distribucija funkcije odluke d^{2008} algoritma ALG_2008 za test slučaj 3.

Tabela 7.2 Greške kod detekcije stridensa za ALG_2008 i ALG_2014

algoritam	Test slučaj 2			Test slučaj 3		
	Promašena detekcija	Nije detektovan	Lažna detekcija	Promašena detekcija	Nije detektovan	Lažna detekcija
ALG_2014	1/16	3/32	1/16	1/16	3/32	0/16
ALG_2008	1/16	5/32	1/16	1/16	9/32	0/16

7.1.4 Rezime

Za razliku od globalne ocene artikulacionih poremećaja koja govori o stepenu razvijenosti i kvalitetu izgovora, artikulacionim testom se analiziraju pojedinačna obeležja koja odgovaraju ispitivanom glasu na osnovu njegove fonetske strukture. U ovom poglavlju analizirana je upotreba algoritama za detekciju stridensa, artikulacionog poremećaja koji se manifestuje kao neprijatan zvuk nalik zvižduku ili kao piskavi ili hrapav zvuk koji se javlja uporedo sa normalnim izgovorom. Predstavljena su dva algoritma za detekciju prisustva stridensa. Prvi je baziran na modelu percepcije zvuka i obuhvata psihoakustičke efekte, dok je drugi baziran na analizi FFT spektra. Poređenjem rezultata dobijenih algoritmima za detekciju stridensa baziranih na FFT spektru (ALG_2008) i auditornom modelu (ALG_2014) vidi se da oba algoritma daju iste rezultate u zonama sa propuštenu i lažnu detekciju. Osnovna razlika je u broju slučajeva koje jedan i drugi algoritam ne mogu klasifikovati. Algoritam ALG_2014 u test slučajevima 2 i 3 ima po 3 stimulusa koje ne može da klasifikuje a ALG_2008 u test slučaju 2 ima 5 a u test slučaju 3 ima 9 stimulusa koje ne može da klasifikuje. Vidi se da je razlika u detekciji algoritama ALG_2014 i ALG_2008 mnogo veća kada su u pitanju subjekti bez stridensa, a sa prisutnom nekom od drugih patologija. Ovo nameće još jedan zaključak da je algoritam ALG_2014 robusniji, odnosno manje osetljiv na prisustvo drugih patologija u odnosu na algoritam ALG_2008. Razlog tome je činjenica da algoritam ALG_2014 bolje podražava auditivnu percepciju stridensa u odnosu na algoritam ALG_2008 te je stoga osetljiv samo na one komponente signala koje doprinose auditivnoj percepciji stridensa. Algoritam za detekciju stridensa ALG_2008 pokazao je dobru usklađenost sa ekspertskim ocenama u slučajevima kada je stridens bio veoma izražen. Zbog rezultata koji su pokazali bolju klasifikaciju stimulusa sa stridensom, algoritam ALG_2014, koji se bazira na auditornom modelu je prihvaćen kao deo sistema u kom se vrši detekcija stridensa. Treba napomenuti da je ovaj algoritam primenljiv i na druge glasove iz grupe frikativa i afrikata, bili oni zvučni ili bezvučni. U tom slučaju potrebno je slobodne parametre algoritma prilagoditi veličinama koje odgovaraju tim glasovima.

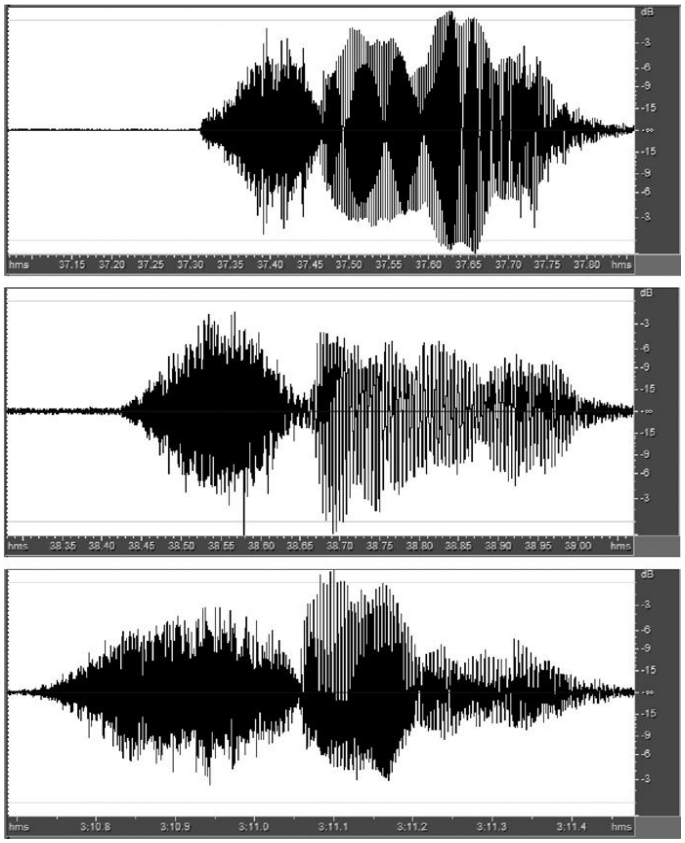
7.2 Detekcija poremećaja trajanja glasa

Trajanje glasa kao akustičko obeležje, bitno je ne samo za razlikovanje tipične i atipične produkcije glasova već i kao distinktivna karakteristika pojedinih fonema. Naime, trajanje fonema može biti produženo do određene granice posle koje se smatra nekim drugim zvukom i na taj način gubi funkcija reči, ili skraćeno do te mere da se percepira kao druga fonema. Postojanje jasnih kriterijuma po pitanju trajanja fonema bitno je kako za razvoj govornih tehnologija (Kato i sar., 2002), tako i za logopedsku teoriju i praksu (Plante i Beeson, 2007). Pod patološkim izgovorom u smislu poremećaja trajanja izgovora foneme, izgovor može biti karakterisan kao produžen ili skraćen. Istraživanja sprovedena u cilju merenja i analize trajanja u domenu patologije glasova, pokazala su da postoje oblasti razdvajanja tipične i atipične produkcije.

7.2.1 Karakteristike trajanja glasa /š/ - pregled

Kako se problem detekcije poremećaja trajanja može rešiti samom segmentacijom fonema u analiziranoj reči, potrebno je bilo dodatno odrediti granice normalnog (tipičnog) izgovora kao i utvrditi da li te apsolutne vrednosti granica fonema mogu biti upotrebljene za distinkciju normalnog i patološkog izgovora ili moraju biti posmatrane relativno u odnosu na ostatak reči. Pored toga, postavlja se i pitanje da li eksperti prilikom ocenjivanja trajanja izgovora uzimaju u obzir samo fonem i njegovo trajanje ili ga posmatraju u širem fonetskom okruženju, odnosno reči.

Na slici 7.17 su, kao primer, data tri reprezentativna uzorka glasa /š/ izgovorenog od strane tri različita govornika. Primer na slici 7.17a predstavlja skraćenu artikulaciju izgovora glasa /š/, primer na slici 7.17b tipičnu, a primer 7.17c produženu artikulaciju. Ocene su date od strane eksperata - logopeda.



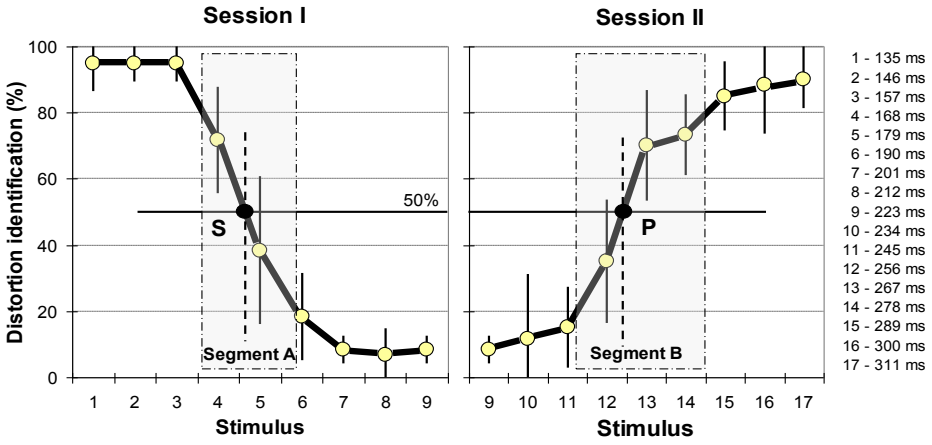
Slika 7.17 Talasni oblici dečjeg izgovora reči /šuma/
(a)skraćena artikulacija /š/ (trajanje /š/ = 155 ms, /uma/ = 385 ms); (b) tipična artikulacija /š/ (trajanje /š/ = 238ms,/uma/ = 397ms); (c) produžena artikulacija /š/ (trajanje /š/ = 345ms, /uma/ = 411ms)

Istraživanja koja su izvedena koristeći produženi, skraćeni i tipični oblik izgovora reči /šuma/ (Jovičić i sar., 2010) pokazala su da ne postoji korelacija između trajanja inicijalnog fonema /š/, i trajanja nastavka reči, /uma/. To dalje znači da se analizirani fonem posmatra izolovano od ostatka reči.

Za implementaciju algoritma potrebno je još utvrditi identifikacione karakteristike trajanja fonema i granice između normalnog i skraćenog, odnosno normalnog i produženog izgovora. Prva istraživanja koja su za cilj imala utvrđivanje ovih granica za glasove srpskog jezika (Jovičić i sar., 2010.) izvedena su variranjem trajanja inicijanog fonema, a na osnovu ocena koje su dali obučeni eksperti – logopedi po pitanju odstupanja trajanja izgovora. Grupu uzoraka činilo je 17 uzoraka govora sintetizovanih tako da je trajanje inicijalnog fonema /š/ u rasponu od 135 ms do 311 ms. Uzorci su formirani na taj način što je najpre uzet uzorak normalnog izgovora reči /šuma/, kod koga je trajanje inicijalnog fonema 223 ms. Zatim je izvršeno skraćivanje i

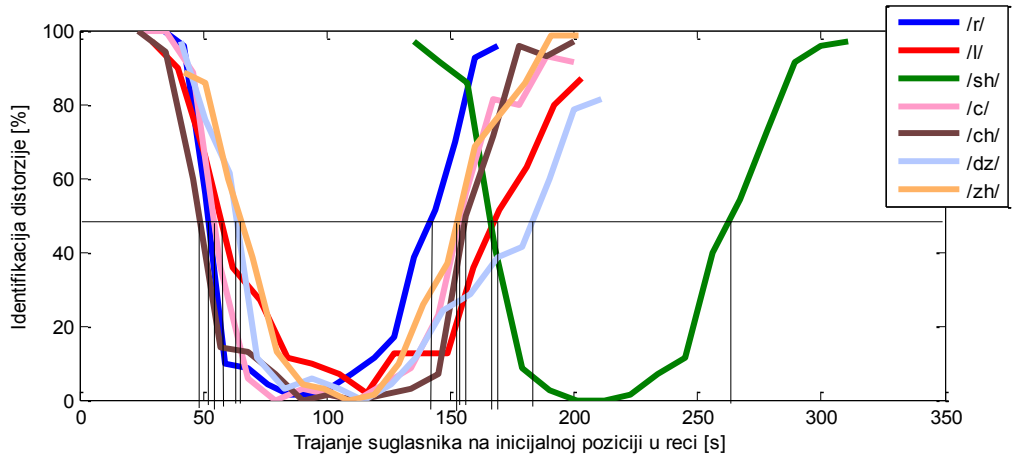
produžavanje trajanja fonema /š/ sa korakom od 11 ms. Ostatak reči /_uma/ je ostao nepromenjen. Na taj način je formirano 8 novih uzoraka skraćivanjem i isto toliko novih uzoraka produživanjem trajanja /š/. Skraćivanje je vršeno uklanjanjem segmenata glasa /š/ dužine 11 ms tako što su za početnu i krajnju tačku segmenta izabrane tačke preseka signala sa nulom na uzlaznim delovima govornog signala. Na taj način je zadržan kontinuitet signala. Produživanje je vršeno ponavljanjem kratkog segmenta frikcije iz središnjeg dela glasa /š/. Pri tome, posebna pažnja je posvećena tome da se zadrži kontinuitet anvelope. Dodatno, izvršena je i perceptivna provera kako bi se identifikovale eventualne frekvencijsko-amplitudske anomalije koje bi mogle nastati nakon ovakve manipulacije signala.

Rezultati u vidu identifikacionih karakteristika trajanja glasa /š/ predstavljeni su na slici 7.18. Stimulusi sa rednim brojevima od 6 do 11 sa verovatnoćom identifikacije ispod 20%, mogu se usvojiti kao tipična trajanja glasa /š/, dok se stimulusi 1, 2 i 3 i 15, 16 i 17 sa verovatnoćama preko 80% mogu identifikovati kao atipična trajanja. To znači da se glas /š/ trajanja dužeg od 285 ms percepira kao devijacija i karakteriše kao produženo trajanje, dok se izgovor istog glasa trajanja kraćeg od 165 ms takođe percepira kao devijacija ali skraćenog trajanja. Oblast između 195 ms i 250 ms obuhvata tipične izgovore glasa /š/ (Jovičić i sar. 2010.). To dalje znači da su ovim skupom uzoraka obuhvaćeni kako tipični izgovori tako i atipični izgovori produženog i skraćenog trajanja. Zona između 165 ms i 195 ms kao i ona između 250 ms i 285 ms, obeležene na slici 7.18 kao oblasti *A* i *B* respektivno, smatraju se zonama neodlučnosti, odnosno, zonama u kojima se ne može doneti pouzdana odluka o prisustvu ili odsustvu patologije.



Slika 7.18 Indentifikacione funkcije za glas /š/.

Dalja istraživanja na polju identifikacionih karakteristika trajanja glasova koja su imala za cilj utvrđivanje oblasti tipičnog i atipičnog trajanja artikulacije podrazumevala su, pored glasa /š/, modifikaciju trajanja izvedenu nad još 6 glasova srpskog jezika (/c/, /č/, /ž/, /dž/, /r/, /l/), koji se nalaze u inicijalnoj poziciji u rečima (/cica/, /čelo/, /džep/, /žaba/, /riba/, /lice/) izgovornih od strane dece uzrasta 10 i 11 godina (Lukić i sar., 2011). Nalik postupku koji je prethodno opisan za varijacije trajanja glasa /š/, navedeni glasovi su izdvojeni nakon čega je vršeno njihovo skraćivanje i produžavanje u 8 koraka fiksnog trajanja, vodeći računa da se održi kontinuitet signala. Ostatak reči (/_ica/, /_elo/, /_ep/, /_aba/, /_iba/, /_ice/) ostao je nepromenjenog trajanja. Na taj način formirano je 17 stimulusa koji se svi međusobno razlikuju po trajanju inicijalnog fonema (8 sintetizovanih skraćenih, 8 produženih i središnji koji nije izmenjen) od kojih su prvi i sedamnaesti stimulus odgovarali atipičnim realzacijama po pitanju trajanja i to prvi skraćenom a sedamnaesti produženom. Sedam logopeda je ocenjivalo navedene stimuluse po pitanju poremećaja trajanja. Na slici 7.19 dat je prikaz rezultata dobijenih za identifikacione funkcije verovatnoće perceptivnog prepoznavanja tipičnog/atipičnog trajanja za glasove, /c/, /č/, /dž/, /š/, /ž/, /r/, /l/. Slično kao i kod glasa /š/ u prethodnom slučaju, za navedenih šest glasova moguće je izdvojiti oblasti tipičnog, produženog i skraćenog izgovora, kao i oblasti neodlučnosti.



Slika 7.19. Indentifikacione funkcije za glasove /c/, /č/, /dž/, /š/, /ž/, /r/, /l/.

7.2.2 Algoritam detekcije poremećaja trajanja glasa

Eksperimentalno utvrđene granice trajanja tipičnog i atipičnog izgovora glasova omogućavaju formiranje algoritma za automatsku detekciju ovih poremećaja. Suštinu ovakvog algoritma čini automatska segmentacija reči koja daje informaciju o trajanju izgovorenog glasa. Blok šema algoritma predstavljena je na slici 7.20.

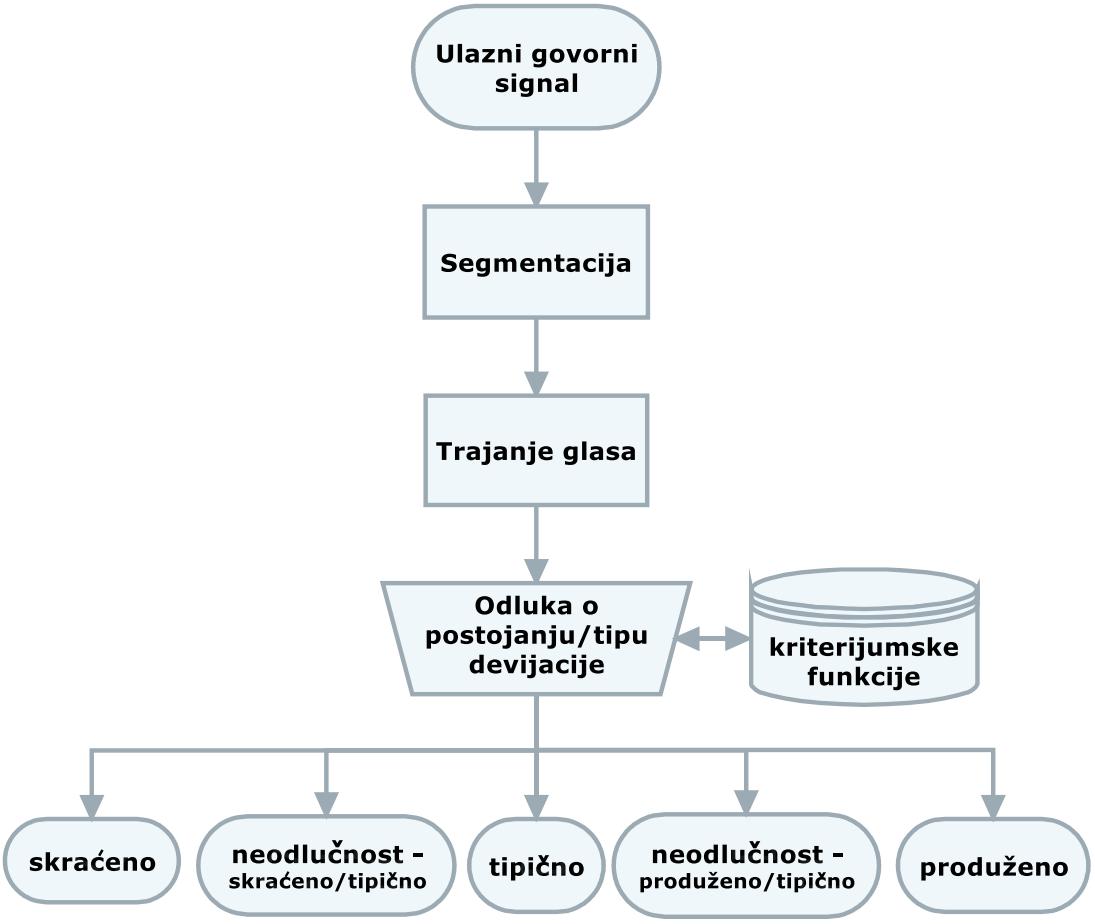
Govorni signal se najpre segmentira. U ovom slučaju za segmentaciju je korišćen algoritam baziran na DTW pristupu, opisan u 5.2.3 dok su za parametrizaciju govornog signala korišćeni MFCC koeficijenti jer su dali bolje rezultate u odnosu na GFCC koeficijente, kao što je pokazano u poglavlju 5.2.4. Na osnovu dobijenih granica fonema, računa se njihovo trajanje. Odluka o postojanju i tipu devijacije u formi patološkog trajanja glasa donosi se prema kriterijumskim funkcijama formiranim na osnovu prethodno opisanih identifikacionih funkcija. U okviru kriterijumskih funkcija, za svaki glas posebno, definisane su granice pet oblasti:

1. Skraćeno trajanje
2. Zona neodlučnosti skraćeno-tipično
3. Tipično trajanje
4. Zona neodlučnosti produženo-tipično
5. Produženo trajanje

Rezultat algoritma je u formi oblasti (jedna od pet gore navedenih) kojoj pripada analizirani glas. Primer kriterijumske funkcije za glas /š/, nastao na osnovu identifikacione funkcije prikazane na slici 7.18, dat je u tabeli 7.3. Slična tabelarna predstava kriterijumske funkcije može se dati i za preostalih šest glasova navedenih u predhodnom odeljku na osnovu njihovih identifikacionih funkcija prikazanih na slici 7.19.

Tabela 7.3 Kriterijumska funkcija za detekciju devijacije trajanja glasa /š/

Kriterijum za trajanje glasa T_g	Oblast
T_g<165 ms	Skraćeno trajanje
165 ms <T_g<195 ms	Zona neodlučnosti skraćeno-tipično
195 ms<T_g<250 ms	Tipično trajanje
250 ms<T_g<285 ms	Zona neodlučnosti produženo-tipično
T_g > 285 ms	Produženo trajanje



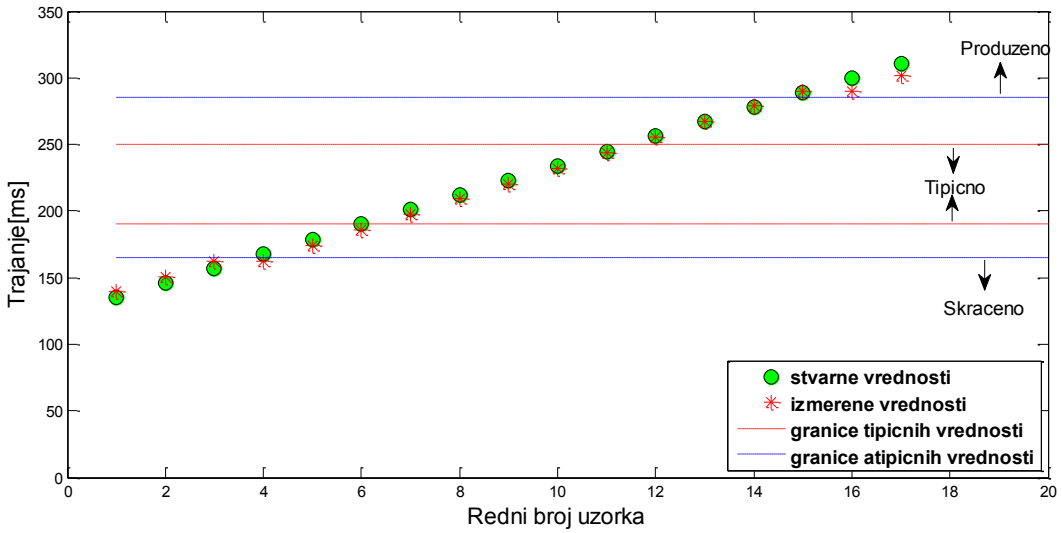
Slika 7.20 Blok šema algoritma detekcije poremećaja trajanja glasa

7.2.3 Rezultati testiranja

Automatska detekcija poremećaja trajanja testirana je na uzorcima reči /šuma/. Testiranje je izvedeno na dve grupe uzoraka – sintetizovanim uzorcima, opisanim u odeljku 7.2.1, i uzorcima iz baze patološkog govora koji su izgovarali odrasli ispitanici, njih 48, a koja je opisana u odeljku 4.5.

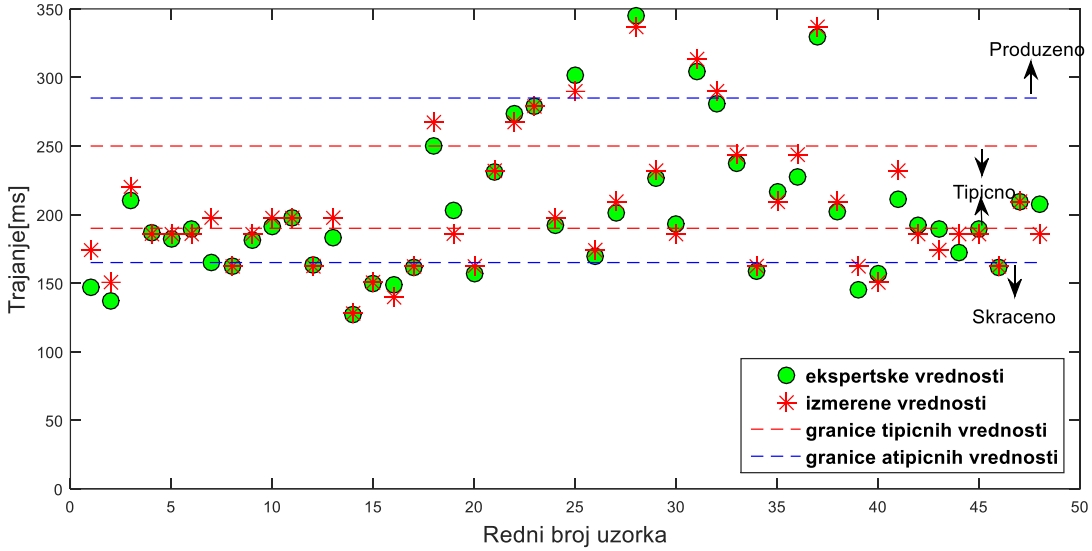
Na slici 7.21 prikazani su rezultati primene opisanog algoritma detekcije poremećaja trajanja glasa na sintetizovanim uzorcima govornog signala. Pored ekspertski izmerenih i automatski detektovanih vrednosti trajanja inicijalnog glasa na slici 7.21 su prikazane i granice tipičnog i atipičnog izgovora, odnosno granice oblasti nesigurnosti (granice određene identifikacionom funkcijom, slika 7.18). Rezultati pokazuju da se vrednosti trajanja inicijalnog glasa poklapaju sa ekspertski izmerenim

vrednostima i da nema uzoraka kod kojih bi bila donesena pogrešna odluka ukoliko bi se koristio automatski pristup. Jedini izuzetak bi eventualno bio uzorak sa rednim brojem 6, trajanja 160 ms, koji se, iako predstavlja granični slučaj, smatra tipičnim izgovorom. Automatskom detekcijom je utvrđeno da bi se u ovom slučaju on svrstao u oblast neodlučnosti, odnosno u oblast za koju nije moguće doneti odluku o prisustvu ili odsustvu devijacije trajanja.



Slika 7.21 *Trajanje glasa /š/ mereno na sintetizovanim uzorcim glasa*

Drugu grupu uzoraka kojom je testiran algoritam činili su uzorci iz baze patološkog govora koje su izgovarali odrasli ispitanici. Testiranje je vršeno koristeći *Leave One Out* (LOO) unakrsnu validaciju (Kohavi, 1995; Cawley i sar., 2003). Rezultati su prikazani na slici 7.22. Utvrđeno je da ni jedan od 48 uzoraka ne bi bio pogrešno dijagnostifikovan koristeći automatsku detekciju. Jedan uzorak skraćenog trajanja i tri uzorka tipičnog trajanja automatskom metodom bi bili smešteni u zonu neodlučnosti, dok bi dva uzorka iz zone neodlučnosti bili smešteni među tipične uzorke. Jedan uzorak koji se nalazi na granici zone neodlučnosti bio bi proglašen produženim izgovorom.



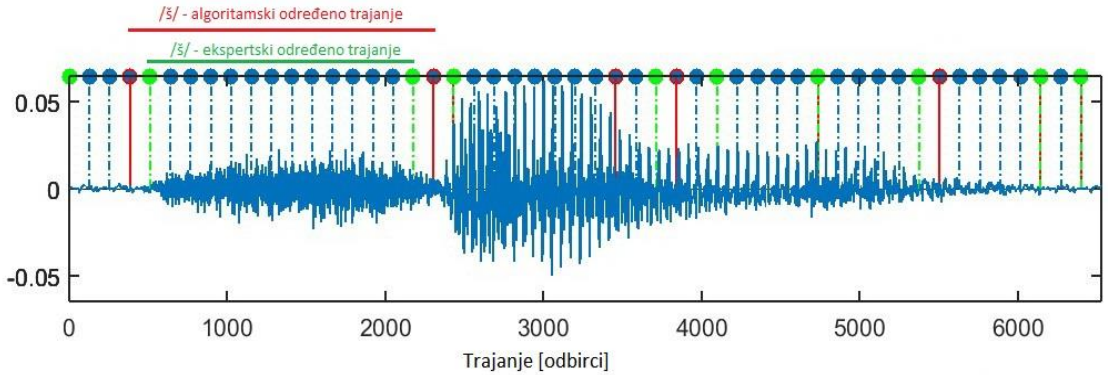
Slika 7.22. Trajanje glasa /š/ mereno na uzorcima iz baze patološkog govora - odrasli

7.2.4 Uticaj algoritma za segmentaciju na detekciju trajanja patološkog glasa

Imajući u vidu da se algoritam za detekciju poremećaja trajanja glasova bazira na algoritmu za segmentaciju, jasno je da će se greške koje nastaju prilikom segmentacije propagirati na ovaj deo sistema. Analizirajući podatke date na sici 7.22 može se zaključiti da će greška segmentacije, ukoliko je po apsolutnoj vrednosti veća od zona neodlučnosti, dovesti do pogrešne i/ili promašene detekcije patološkog, odnosno normalnog izgovora. Neki uzorci koji prema ekspertskoj oceni pripadaju oblastima skraćenog ili produženog izgovora, automatskom detekcijom mogu biti smešteni u zonu normalnog izgovora, i obratno, ukoliko je greška segmentacije veća od zona neodlučnosti. Mogu se izgvojiti dva karakteristična slučaja u kojima greška segmentacije najviše utiče na izračunavanje trajanja glasa.

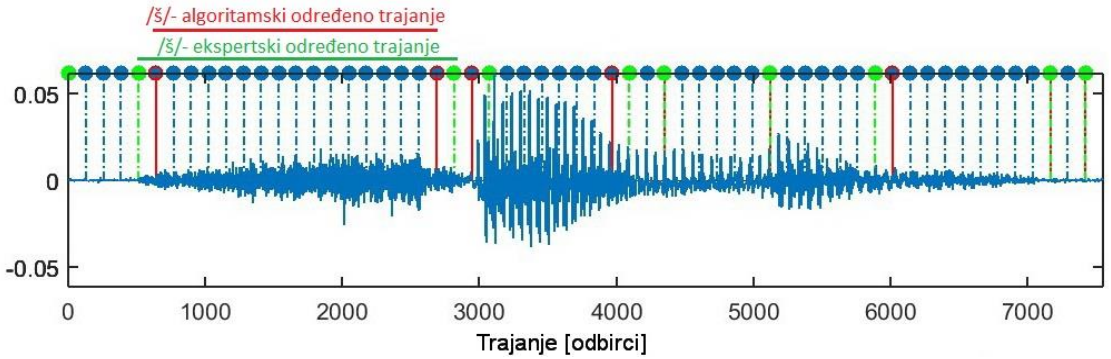
U prvom slučaju, algoritam prvu granicu glasa detektuje pre a drugu posle ekspertski određene granice, što dovodi do toga da je algoritamski dobijeno trajanje veće od ekspertski određenenog. Jedan takav slučaj prikazan je na slici 7.23 i odgovara

uzorku sa rednim brojem 1 sa slike 7.22. To dalje znači da uzorci skraćenog i normalnog trajanja mogu biti detektovani kao normalni, odnosno produženi.



Slika 7.23 Ekspertski i algoritamski određene granice glasa /š/ u reči šuma. Crvenom su date algoritamski određene granice a zelenom ekspertski određene granice.

U drugom slučaju, algoritamskim postupkom prva granica glasa detektuje se posle a druga pre ekspertski određenih granica, što dovodi do toga da algoritamski dobijeno trajanje ima manju vrednost od ekspertskog. Primer za ovaj slučaj prikazan je na slici 7.24 i odgovara uzorku sa rednim brojem 48 sa slike 7.22. Ovo za posledicu može imati pogrešnu detekciju, odnosno, uzorci produženog i normalnog trajanja mogu biti detektovani kao normalni, odnosno skraćeni.



Slika 7.24 Ekspertski i algoritamski određene granice glasa /š/ u reči šuma. Crvenom su date algoritamski određene granice a zelenom ekspertski određene granice.

Kako su veličine zona neodlučnosti 30 ms (tipično-skraćeno) i 35 ms (tipično-produženo), u najgorem slučaju, dovoljno je da greška segmentacije bude veća od ovih

vrednosti kako bi se ispoljila greška kod algoritma detekcije poremećaja trajanja glasa. Ukoliko se u razmatranje uzme strožiji uslov za odstupanje algoritma za segmentaciju (poglavlje 5.2.4.1), od 2 rama analize, u najgorem slučaju, greška kod algoritamske detekcije će biti 4 rama analize, odnosno 45.6 ms, što jeste veće od veličina zona nesigurnosti. Kako se greška segmentacije za korišćeni algoritam i odstupanje veće od 2 rama analize javlja u 4.46% (na nivou granica fonema, Tabela 5.6), to će se odraziti na algoritam za detekciju poremećaja trajanja glasa pa će se greška veća ili jednaka 4 bloka obrade javiti u 2% slučajeva. Testiranjem na opisanoj grupi uzoraka iz baze patološkog govora odraslih govornika ustanovljeno je da nije postojao slučaj u kom je odstupanje granica bilo veće od 2 rama analize, odnosno slučaja lažne ili promašene detekcije.

7.2.5 Rezime

U ovom delu predstavljen je algoritam za detekciju poremećaja trajanja glasova. Algoritam je baziran na automatskoj segmentaciji glasova kojom se izdvajaju granice glasa od interesa i vrši merenje njegovog trajanja. Koristeći identifikacione funkcije kojima su definisane granice normalnog u patološkog trajanja glasova, algoritam donosi odluku o pripadnosti ispitivanog glasa jednoj od pet oblasti koje označavaju: skraćeno trajanje, neodlučnost skraćeno-tipično trajanje, tipično trajanje, neodlučnost produženo-tipično trajanje, produženo trajanje. Rezultati testiranja predloženog algoritma na sintetizovanim uzorcima pokazali su da se ekspertski utvrđene ocene o poremećaju trajanja glasa poklapaju u potpunosti sa algoritamskim ocenama. Pored sintetizovanih uzoraka, i testiranje na uzorcima iz baze patološkog govora pokazalo je da se po pitanju patološkog trajanja glasa ekspertske i algoritamske ocene podudaraju. Kako je greška algoritma direktno zavisna od greške segmentacije, analizom je ustanovljeno da bi do pogrešne detekcije došlo u 2% slučajeva.

8 Zaključak

8.1 Pregled rezultata

Prateći utvrđenu (u nauci i praksi) metodologiju procene patologije govora prikazan je razvoj sistema za detekciju patološkog govora od inicijalnih teorijskih osnova, preko analize uzoraka iz prakse, postavljanja različitih modela, testiranja i analize dobijenih rezultata.

Postavljanje modela segmentacije glasova prilagođenog patološkom izgovoru može se sa visokom tačnošću izvesti kombinacijom dinamičkog vremenskog usklađivanja i metode *k* najbližih suseda. Od ispitivanih parametrizacija (MFCC i GFCC koeficijenata), ona kod koje je govor predstavljen MFCC koeficijentima pokazala se kao bolji pristup. Prosečna greška segmentacije u slučaju korišćenja MFCC koeficijenata i dinamičkog vremenskog usklađivanja bila je 2.26%.

Pokazano je da se za formiranje modela za procenu postojanja artikulacionih odstupanja glasova u izgovoru sa velikom pouzdanošću mogu koristiti neuralne mreže. Analiza različitih topologija neuralnih mreža i ulaznih parametara pokazala je da izdvajanje obeležja govornog signala pomoću MFCC koeficijenta kao ulaznih parametara za višeslojni perceptron daje dobre rezultate kod detekcije globalne ocene patologije glasa. Dalje usložnjavanje topologije dovelo je do upotrebe ansambla neuralnih mreža a testiranje na grupi od 200 dece pokazalo je da se najbolji rezultati dobijaju upravo u ovom slučaju. Adekvatnim izborom broja "eksperata" u ansamblu tačnost prepoznavanja patologije pojedinačnog glasa bila je i do 98%. Što se ulaznih parametara tiče, najbolje se pokazala kombinacija MFCC koeficijenata i karakteristika trajanja i energije analizirane reči i inicijalnog fonema.

Razvoj pojedinačnih modela za detekciju specifičnih oblika odstupanja akustičkih obeležja u vremenskom, amplitudskom, spektralnom i parametarskom domenu je veoma široka oblast istraživanja imajući u vidu sve moguće varijante

devijacije izgovora i njihove patološke realizacije. U ovoj disertaciji posmatrana su odstupanja izgovornog glasa po pitanju pojave stridensa i devijacije trajanja.

Automatska detekcija stridensa ispitivana je pomoću dva algoritma: algoritmom za detekciju stridensa baziranih na FFT spektru (ALG_2008) i auditornom modelu (ALG_2014). Analizirajući vrednosti dobijene za oba algoritma, može se zaključiti da oni daju iste rezultate u zonama sa propuštenom i lažnom detekcijom i da je osnovna razlika u broju uzoraka koje i jedan i drugi algoritam ne mogu klasifikovati. Algoritam ALG_2014 je imao manje stimusa koji su neklasifikovani. Specijalno, razlika u klasifikaciji između ALG_2014 i ALG_2008 bila je mnogo veća kada su u pitanju subjekti bez stridensa, a sa nekom od drugih patologija. ALG_2014 se pokazao kao robusniji, odnosno manje osetljiv na prisustvo drugih patologija.

Detekcija poremećaja trajanja glasa bazirana na algoritnu za segmentaciju pokazala je dobre rezultate kako na sintetizovanim uzorcima tako i na realnim uzorcima iz baze patološkog govora. Sintetizovanim uzorcima je bila pokrivena oblast trajanja glasa /š/, odnosno daljim skraćivanjem glas /š/ bi prešao u glas /č/ a daljim produživanjem bi se izgubila suštinska karakteristika glasa, pa se može reći da je algoritmom pouzdano detektovan ceo opseg trajanja. Rezultati dobijeni za stimuluse koji nisu sintetizovani i koji su deo baze patološkog govora pokazali su da 7 uzoraka od 48 testiranih (15%) nije bilo smešteno u zonu kojoj pripada, međutim ni jedan od njih nije pogrešno dijagnostifikovan, odnosno ni jedan od uzoraka tipičnog trajanja nije proglašen patološkim ili obnuto. Posmatrajući efekat koji algoritam za segmentaciju ima na određivanje trajanja glasa po pitanju veličine greške, pokazano je da bi u 2% slučajeva došlo do pogrešne detekcije po pitanju postojanja ove vrste patologije govora.

8.2 Doprinos disertacije

Najznačajniji doprinos ove disertacije je svakako u tome što je pokazano da se na osnovu istraživanja na polju distinktivnih karakteristika normalnog i patološkog govora može razviti sistem za automatsku procenu artikulacionih poremećaja baziran na metodama razvijenim u oblasti obrade govornog signala. Sistem u svim elementima u potpunosti podržava metodologiju koja se koristi u logopedskoj praksi što ga čini pogodnim za uporedno testiranje i, što je najbitnije, korišćenje od strane terapeuta. Konkretni rezultati proistekli iz ove disertacije obuhvataju, na prvom mestu, formiranje

jedinstvenog modela za prepoznavanje artikulaciono - akustičkih odstupanja glasova u patološkom izgovoru. Ovakav sistem se može koristiti samostalno, ili u okviru nekih drugih sistema gde je potrebna relativno brza procena kvaliteta izgovora glasova. Postavljanje modela segmentacije glasova sa izraženim odstupanjima pojedinih akustičkih obeležja takođe je jedan od rezultata koji su proistekli iz ove disertacije. Visoka tačnost modela segmentacije i jednostavnost implementacije omogućava njegovu primenu i u drugim sistemima gde je potrebno izvršiti segmentaciju reči a gde je poznat govorni sadržaj. Pored navedenih, i razvoj pojedinih modela za detekciju specifičnih oblika odstupanja akustičkih obeležja u vremenskom, spektralnom i parametarskom domenu predstavlja jedan od konkretnih rezultata proisteklih iz ove disertacije. Automatizacija detekcije govorne patologije do nivoa konkretne patološke pojave posebno je značajna sa aspekta logopedske prakse jer praktično dopinosi formiranju etalona za klasifikaciju tipičnog i patološkog glasa.

Predstavljena istraživanja i njihovi rezultati pokazala su da se korišćenjem pristupa priznatih u logopedskoj praksi, a pomoću tehnika obrade govornog signala, može zaokružiti sistem za automatsku detekciju patoloških glasova srpskog jezika. Ocene koje se na ovaj način dobijaju sa visokom tačnošću se podudaraju sa ocenama dobijenim od strane obučenih eksperata - logopeda. Visoka tačnost predloženog sistema govori u prilog tome da se on može koristi za skrining testiranja većih razmera u cilju ranog otkrivanja i prepoznavanja poremećaja govora, kao i podrška u logopedskoj terapiji i lečenju, evaluaciji terapijskih strategija, praćenju logopedskog tretmana i kliničkim ispitivanjima. Realizacija ovakvog sistema može dati značajan doprinos u okviru e-medicine tehnologije.

8.3 Mogućnost daljih istraživanja

U ovoj disertaciji prikazano je rešenje sistema za detekciju patologije govora. Najperspektivniji deo sistema što se daljih istraživanja tiče jeste modul za detekciju artikulacionih poremećaja koji se bazira na analitičkoj oceni patologije izgovora. Prikazano rešenje obuhvata dve manifestacije patološkog govora – stridens i izmenjeno trajanje glasa, međutim, modularna struktura sistema omogućava njegovu nadogradnju i proširenje. Inicijalna istraživanja bazirana na analizi uporednih karakteristika normalnog i patološkog izgovora kod poremećaja intenziteta glasova i određenih tipova

sigmatizama (interdentalnog i adentalnog) dala su rezultate koji mogu poslužiti kao osnova za dalja istraživanja i razvoj algoritama za njihovu automatsku detekciju. Za pomenuta odstupanja moguće je definisati kriterijume u akustičkom domenu po kojima bi bila moguća diferencijacija tipičnog i atipičnog izgovora.

Analiza karakteristika intenziteta tipičnih i patološki izgovorenih glasova (Punišić i sar., 2011b, Punišić, 2012) dala je distribucije intenziteta na osnovu kojih je moguće definisati oblasti u kojima je intenzitet tipičan i one u kojima je jak ili slab, odnosno patološki. Ovakve kriterijumske funkcije moguće je direktno primeniti u okviru predstavljenog sistema za automatsku detekciju patološkog izgovora.

Analiza akustičkih korelata odstupanja u spektralnom domenu, pokazala je da se odstupanja po tipu interdentalnog i adentalnog sigmatizma mogu prepoznati u spektralnom domenu (Vojnović i Punišić, 2010; Vojnović i Punišić, 2011; Punišić, 2012). Za pomenuta odstupanja moguće je definisati kriterijume diferencijacije tipičnog i atipičnog izgovora na osnovu energetskih odnosa pojedinih podopsega u govornom signalu.

Uočavanje i dokazivanje karakteristika određene patologije i njena translacija u domen parametara zahteva sveobuhvatan multidisciplinarni pristup i neizostavnu analizu velikog broja uzoraka, kako normalnog tako i patološkog govora. Da bi sistem obuhvatio sve oblike patološkog izgovora potrebna su na prvom mestu dalja detaljna fundamentalna istraživanja koja bi u parametarskom domenu dala oblasti razdvajanja patološkog i normalnog glasa. Predloženi algoritmi mogli bi se onda jednostavno prekonfigurisati i koristiti za druge glasove proširujući obučavajuće skupove i podešavajući algoritamske parametre na za to odgovarajuće vrednosti. U tom smislu opisani sistem predstavlja dobru osnovu i ima mogućnost daljeg razvoja.

Literatura

Akbari, A., Arjmandi, M.K. (2014). An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. Biomedical Signal Processing and Control 10 (2014) 209–223

Alonso, J.B., de Leon, J., Alonso I., and Ferrer, M.A. (2001). Automatic detection of pathologies in the voice by HOS based parameters. EURASIP Journal on Advances in Signal Processing, vol. 4, pp. 275-84, 2001.

Alonso, J.B., Díaz de María, F., Travieso, C.M., and Ferrer, M.A. (2005). Using nonlinear features for voice disorder detection. Proceedings of the 3rd International Conference on Non-Linear Speech Processing NOLISP, Barcelona, Spain, pp. 94-106, 2005.

Andre-Obrecht, R. (1988). A new Statistical Approach for the Automatic Segmentation of the Continous Speech Signals. IEEE Trans. on Acoust. Speech Signal Proceeding, ASSP-36, No.1, pp. 29-40, January 1988.

Appel, U., Brandt, A.V. (1983). Adaptive Sequental Segmentation of piecewise Stationary Time Series. Information Science, Vol. 29, No.1, pp. 27-56, 1983.

Arias-Londoño, J.D., Godino-Llorente, J.I., Sáenz-Lechón, N., Osma-Ruíz V., and Castellanos-Domínguez, G. (2011). Automatic detection of pathological voices using complexity measures, noise parameters and mel-cepstral coefficients. IEEE Transactions on Biomedical Engineering, vol. 58(2), pp. 370-9, 2011.

Azarshid, F., Perennou, G., and Andre-Obrecht, R. (1993). A Segmental Approach Versus a Centisecond One for Automatic Phonetic Time-Alignment. Third European Conference on Speech Communication and Technology, EUROSPEECH' 93 Berlin, Germany, September 22-25, pp. 657-660, 1993.

Benselama, Z., Guerti, M., and Bencherif, M. (2007). Arabic speech pathology therapy computer aided system. Journal of Computer Science, vol. 3, no. 9, pp. 685–692, 2007.

Bilibajkić, R., Šarić, Z., Jovičić, S.T. (2007). Segmentacija reči postupkom najbližeg uzorka za potrebe analize poremećaja izgovora fonema. 51. Konferencija ETRAN Herceg Novi.

Bilibajkić, R., Šarić, Z., Jovičić, S.T. (2010). Segmentacija reči za potrebe dijagnostike patologije govora primenom auditornog modela. 54. Konferencija ETRAN Donji Milanovac.

Bilibajkic, R. (2011). Segmentacija reči na bazi MFCC i GFCC spektralnih modela. Magistarski rad, Elektrotehnički fakultet, Univerzitet u Beogradu.

Bilibajkić, R., Šarić, Z., Jovičić, S.T. (2011). Auditory-model based speech segmentation on subphonemic segments. Forum Acusticum 2011. Proceedings 2011; Aalborg, Denmark, ISBN: 978-84-694-1520-7, pp. 55-60, Aalborg, Danska.

Bilibajkić, R., Šarić, Z., Punišić, S., Subotić, M., Jovičić, S. (2012). Detekcija stridensa u patološkom izgovoru primenom auditornog modela. Zbornik radova sa simpozijuma Digitalna obrada govora i slike, DOGS 2012, 4-7. oktobar 2012, Kovačica.

Bilibajkić, R., Šarić, Z., Punišić, S., Subotić, M., Jovičić, S. (2013). Automatic detection of stridence in speech using the auditory model. Proceedings of 4th Iternational conference on Fundamental and Applied Aspects of Speech and Language, pp. 230 -239, Belgrade, October 2013.

Bilibajkić, R., Subotić, M., Furundžić, D. (2014). Primena neuralnih mreža u detekciji patološkog izgovora srpskih glasova. ZBORNIK RADOVA XXII TELEKOMUNIKACIONI FORUM TELFOR 2014, Izdavači: Društvo za telekomunikacije - Beograd, Akademska misao - Beograd, 25-27. novembar, Beograd, Srbija. ISBN: 978-1-4799-6190-0, pp. 873-876.

Bilibajkić, R., Šarić, Z., Jovičić, S., Punišić, S., Subotić, M. (2015). Automatic detection of stridence in speech using the auditory model. Computer Speech & Language, Advance online publication. DOI: http://dx.doi.org/10.1016/j.csl.2015.08.006. To appear in: Computer Speech & Language, March 2016, Volume 36, pp.122–135.

Black, A.W., Kominek, J., Bennett, C. (2003). Evaluating and Correcting Phoneme Segmentation for Unit Selection Synthesis. In: Proceeding Eurospeech, Geneva, Switzerland, pp. 313–316 (2003).

Boersma, P., and Weenink, D. (2010). Praat: doing phonetics by computer (version 5.1.29). Available from http://www.praat.org/. [Computer program].

Boyanov, B., Hadjitodorov, S. (1997). Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases. IEEE Engineering in Medicine &Biology Magazine 16 (4) (1997) pp.74–82.

Bunnell, H.T., Debra, M.Y., and Polikoff, J.B. (2000a). Using Markov models to assess articulation errors in young children. In The Journal Of The Acoustical Society of America, Vol. 107, Issue 5, p. 2093, (2000a).

Bunnell, H.T., Yarrington, D.M., and Polikoff, J.B. (2000b). STAR: Articulation training for young children. In Proceedings of the ICSLP 2000, vol. 4, pp. 85-88.

Burg, J. P. (1967). Maximum Entropy Spectral Analysis. Presented at the 37th Annual International Meeting, Soc. of Explor. Geophysics, Oklahoma, Oct. 1967.

Burg, J. P. (1975). Maximum Entropy Spectral Analysis. PhD thesis, Stanford University, 1975. Available from http://sepwww.stanford.edu/theses/sep06/

Cairns, D.A., Hansen, J.H.L., and Riski, J.E. (1994). Detection of hypernasal speech using a nonlinear operator. Proceedings of IEEE Conference on Engineering in Medicine and Biology Society, pp. 253-4, 1994.

Cairns, D.A., Hansen, J.H.L., and Riski, J.E. (1996). A noninvasive technique for detecting hypernasal speech using a nonlinear operator. IEEE Transactions on Biomedical Engineering, vol. 43, no. 1, pp. 35-45, 1996.

Cawley, G.C., Talbot, N.L.C. (2003). Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers. Pattern Recognition, 04/2003; 36(11):2585-2592.

Chomsky, N., and Halle, M. (1968). The Sound Pattern of English. (Harper and Row, New York).

Chou, F.C., Tseng, C.Y., and L.S. (1998). Automatic segmental and prosodic labeling of Mandarin speech database. Proceeding of the 5fth International Conference on Spoken Language Sydney, Australia, November 30 - December 4, 1998. Processing. www.isca-speech.org/arhive.

Cosi, P., Falavigna, D., and Omologo, M. (1991). A Preliminary Statistical Evaluation of Manual and Automatic Segmentation Discrepancies. Proceedings of EUROSPEECH-1991, 2ndEuropean Conference on Speech Tehnology, Genova, 24-26 September, 1991, pp. 693-696.

Danubianu, M., Pentiuc, S.G., Schipor, O., Nestor, M., Ungurean, I., Schipor, D.M. (2009). TERAPERS - Intelligent Solution for Personalized Therapy of Speech Disorders. International Journal on Advances in Life Science 1(1), 26–35.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation. I. modulation detection and masking with narrowband carriers. J. Acoust. Soc. Am., 102(5):2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. II. spectral and temporal integration in modulation detection. J. Acoust. Soc. Am., 102(5):2906–2919.

Dau, T. (2009). Auditory processing models. In Havelock, D., Kuwano, S., and Vorl¨ander, M., editors, Handbook of Signal Processing in Acoustics, pp. 175–196. Springer New York, NY, USA.

Davis, S., Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366.

de Krom, G.A. (1993). Cepstrum-based technique for determining a harmonics-to-noise rati oin speech signals. Journal of Speech and Hearing Research 36(2) (1993): 254–266.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. Journal of Speech and Hearing Research 38 (August (4)) (1995) :794–811.

Delgado, E., Sepúlveda, F.A., Röthlisberger, S., and Castellanos, G. (2011). The Rademacher Complexity Model over acoustic features for improving robustness in hypernasal speech detection. Computers and Simulation in Modern Science, vol. 5, pp. 130-5, WSEAS Press, University of Cambridge, UK, 2011.

Deng, L., Geisler, C.D., and Greenberg, S.(1988). A composite model of the auditory periphery for the processing of speech. Journal of Phonetics 16, 93-108, 1988.

Dibazar, A., Narayanan, S., Berger, T.W. (2002). Feature analysis for automatic detection of pathological speech, in: Proceedings of the Second Joint EMBS/BMES Conference, vol.1, Houston, TX, USA, 2002.

Dibazar, A., Berger, T., Narayanan, S. (2006). Pathological voice assessment, in: Proceedings ofthe28 th Annual International Conference of the IEEE EMBS, New York , NY, USA, 2006.

Escartın, A., Saz, O., Vaquero, C., Rodrı´guez, W.R., Lleida, E. (2008). Comunica Framework Website. <http://www.vocaliza.es>.

Fabrice, M., Deroo, O., and Dutoit, T. (1998). Phonetic alignment: Speech synthesis based vs. hybrid hmm/ann. Proceedings of the ICSLP, 1998.

Feijoo, S., Hernandez-Espinosa, C. (1990). Short-term stability measures for the evaluation of vocalquality. Journal of Speech and Hearing Research 33(1990)324–334.

Feldbauer, C., Kubin, G., Kleijn, W.B. (2005). Anthropomorphic Coding of Speech and Audio: A Model Inversion Approach, EURASIP Journal on Applied Signal Processing 2005:9, 1334-1349.

Fonseca, E.S., Guido, R.C., Scalassara, P.R., Maciela, C.D., Pereira, J.C. (2007). Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. Computers in Biology and Medicine 37(4)(2007)571–578.

Fraile, R., Saenz-Lechon, N., Godino-Llorente, J.I., Osma-Ruiz, V., Fredouille, C. (2009). Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficients parameters and differentiation of patients by sex, Folia Phoniatrica et Logopaedica 61(3)(2009) 146–152.

Fredouille, C., Pouchoulin, G. (2011). Automatic detection of abnormal zones in pathological speech. In International Congress of Phonetic Sciences (ICPHs'11), Hong Kong.

Fujimura, O., and Lindqvist, J. (1971). Sweep-tone measurements of the vocal tract characteristics. Journal Acoustical Soc. Am., vol. 49(2), pp. 541-58, 1971.

Furundžić, D., Subotić, M., Pantelić, S. (2006). Ocena poremećaja govora na nivou fonema primenom neuralnih mreža. Proceedings of Conference DOGS, Vršac, 2006. pp. 10-13.

Furundžić, D., Subotić, M., Pantelić, S. (2007). Primena neuronskih mreža u klasifikaciji poremećaja izgovora frikativa. Proceedings of 51th Conference ETRAN, 2007.

Furundzić, D., Jovičić, S. T., Subotić, M. Z., Punišić, S. (2012). Acoustic Features Determination for Regularity Articulation Quantification of Serbian Fricatives. 11. symposium on neural network applications in electrical engineering (NEUREL 2012).

Furundžić, D., Jovičić, S.T., Subotić, M., Punišić, S. (2013). Optimization process classification of articulation disorders, VERBAL COMMUNICATION QUALITY Interdisciplinary Research II Edited by Jovičić, S.T., Subotić, M., Sovilj, M. LAAC & IEPSP, Belgrade 2013.

Gavat, I., Grigore, O., Velican, V. (2011). Impaired Speech Recognition. Case Study: Recognition of Initial 'r' Consonant in Rhotacsim Affected Pronunciations, in Proceedings of the 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2011, Brasov,Romania, pp.1-6.

Gelzinis, A., Verikas, A., Bacauskiene, M. (2008). Automated speech analysis applied to laryngeal disease categorization. Computer Methods and Programs in Biomedicine 91(1) (2008)36–47.

Ghitza, O. (1986). Auditory nerve representation as a front-end for speech recognition in anoisy environment. Computer Speech and Language, vol. 1, pp. 109–130, 1986.

Giovanni, A., Ouaknine, M., and Triglia, J.M. (1999). Determination of largest Lyapunov exponents of vocal signal: application tounilateral laryngeal paralysis. J. voice, vol. 13(3), pp. 341-54, 1999.

Glasberg, B.R., and Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. Hearing Research, 47: 103-138.

Godino-Llorente, J.I., Gomez-Vilda, P., Blanco-Velasco, M. (2006). Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. IEEE Transactions on Biomedical Engineering 53 (10)(2006)1943–1953.

Godino-Llorentea, J.I., Frailea, R., Sáenz-Lechóna, N., Osma-Ruiza, V., Gómez-Vildab, P. (2009). Automatic detection of voice impairments from text-dependent running speech. Biomedical Signal Processing and Control Volume 4, Issue 3, July 2009, Pages 176–182.

Golubović, S. (1997). Klinička logopedija I. Defektološki fakultet, Beograd.

Grigore, O., Grigore, C., Velican, V. (2010). Intelligent System for Impaired Speech Evaluation. In Proceedings of the International Conference on Circuits, Systems, Signals, 10/2010, pp. 365-368.

Hadjitodorov, S., Mitev, P. (2002). A computer system for acoustic analysis of pathological voices and laryngeal disease screening. Medical Engineering & Physics 24(6)(2002)419–429.

Hammaberg, B., Fritzell, B., Gauffin, J., Sundberg, J. (1980). Perceptual and acoustic correlates of abnormal voice qualities. Acta Otolaryngologica 90 (1980) 441–451.

Hansen, L.K., Salamon, P. (1990). Neural network ensembles. IEEE Transactions Pattern Analysis and Machine Intelligence 12(10), 993-1001.

Hohmann, V. (2002). Frequency analysis and synthesis using a Gammatone filterbank. Acta Acustica United with Acustica, Vol. 88 (2002) 433 – 442.

Holdsworth, J., Nimmo-Smith, I., Patterson, R., and Rice, P. (1988). Implementing a gammatone filter bank, Annex C of the SVOS Final Report: Part A: The Auditory Filterbank , 1988.

Holt, L.L., Lotto, A. (2010). Speech perception as categorization Attention. Perception, & Psychophysics, 2010, 72 (5), 1218-1227.

Honova, J., Jindra, P., Pešák, J. (2003). Analysis of articulation of fricative prealveolar sibilant “s” in control population. Biomedical Papers 147(2), 239–242 (2003).

Huggins-Daines, D., Rudnicky, A.I. (2006). A Constrained Baum-Welch Algorithm for Improved and Efficient Training. In: Proc. Interspeech 2006s-9th International Conference on Spoken Language Processing, Pittsburgh, USA (2006).

Irino, T, Patterson, R.D. (1997). A time-domain. level-dependent auditory filter: the gammachirp. J. Acoust. Soc. Am. 101: 412-419.

Irino, T., Patterson, R.D. (2001). A compressive gammachirp auditory filter for both physiological and psychophysical data. J. Acoust. Soc. Am. 109(5 Pt 1):2008-22.

Itakura, F. (1975). Minimum Prediction Residual Principe Applied to Speech Recognition. IEEE Trans. Acoust. Speech Siganal Processing, ASSP-23, No.1, pp. 67-71, February 1975.

Jepsen, M.L., Ewert, S.D., and Dau, T. (2008). A computational model of human auditory signal processing and perception. J. Acoust. Soc. Am., 124(1):422–438.

Jiang, J., Zhang, Y., and Mcgilligan, C. (2006). Chaos in voice, from modeling to measurement. J. Voice, vol. 20(1), pp. 2-17, 2006.

Jokic, I.D., Jokic, S.D., Peric, Z.H., Delic, V.D. (2014). Towards a Small Intra-Speaker Variability Models. Electronics and Electrical Engineering, Vol. 20, No. 6, pp. 100-103, 2014.

Jokić, I., Bilibajkić, R., Šarić, Z., Jovičić, S. (2014). Prepoznavanje stridensa na bazi modela zasnovanog na višedimenzionalnoj Gausovoj raspodeli. ZBORNIK RADOVA DOGS 2014, X konferencija digitalna obrada govora i slike. Izdavači: Fakultet Tehničkih Nauka - Novi Sad, Elektrotehnički Fakultet - Beograd, Elektronski Fakultet - Niš, 5-9 Oktobar, 2014, Novi Sad, Srbija. ISBN:978-86-7892-633-4, pp 19-22.

Jovičić, S.T. (1999). Govorna komunikacija: fiziologija, psihoakustika i percepcija. Nauka, Beograd.

Jovicic, S., Punisic S., Saric Z. (2008). Time-frequency detection of stridence in fricatives and affricates. Acoustics'08, Paris, 5137-5141.

Jovicic, S., Kasic, Z., Punisic S., (2010). Production and perception of distortion in word-initial friction duration, Journal of communication disorders, 2010, vol. 43 br. 5, 335-346.

Kaiser, J.F. (1990). On a simple algorithm to calculate the “energy” of a signal. In: 1990 International Conference on Acoustics, Speech, and Signal Processing. ICASSP-1990. New York: IEEE, 1990: 381–4.

Karjalainen, M., Altosaar, T., and Huttunen, M. (1998). An e_cient labelling tool for the QUICKSIG speech database. Proceedings of the International Conference on Spoken Language Processing (1998): 1535-1538.

Kasuya, H., Ogawa, S., Mashima, K., Ebihara, S. (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. Journal of the Acoustical Society of America 80 (5) (1986) 1329–1334.

Kato, H., Tsuzaki, M., Sagisaka, Y. (2002). Effects of phonetic quality and duration on perceptual acceptability of temporal changes in speech. Journal of the Acoustical Society of America, 111(1 Pt 1), 387–400.

Kent, R.D. (1996). Hearing and Believing: Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders. American Journal of Speech-Language Pathology 5, pp. 7-23.

Kinnunen, T., Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. Speech Communication 52, pp. 12-40, 2010.

Klingholtz, F., Martin, F. (1987). The measurement of the signal-to-noiseratio (SNR) in continuous speech. Speech Communication 6 (1) (1987)15–26.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings IJCAI-95, Montreal, Que. (Morgan Kaufmann, Los Altos, CA, 1995), 1137-l 143.

Kostić, Đ., Nestorović, M., Kalić, D. (1964). Akustička fonetika srpskohrvatskog jezika. Beograd: IEFPG, 1964.

Kostić, Đ., Vladisavljević, S., Popović, M. (1983): Testovi za ispitivanje govora i jezika. Zavod za udžbenike i nastavna sredstva, Beograd.

Kvale, K. (1993). Segmentation and Labeling of Speech. PhD Dissertation. The Norwegian Institute of Technology, 1993.

Lee, C., Hyun, D., Choi, E., Go, J., and Lee, C. (2003). Optimizing Feature Extraction for Speech Recognition. IEEE Transactions on Speech and Audio Processing, Vol. 11, No. 1, pp. 80-87, January 2003.

Linder, R., Albers, A.E., Hess, M., Poppl, S.J., Schonweiler, R. (2008). Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features. Journal of Voice 22(2)(2008)155–163.

Little, M., Costello, D., and Harries, M. (2011). Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures. Journal of Voice, vol. 25(1), pp. 21-31, 2011.

Livingston, K.R., Andrews, J.K., Harnad, S. (1998). Categorical Perception Effects Induced by Category Learning. Journal of Experimenntal Psychology: Learning, Memory, and Cognition ,24, 732-753.

Lleida, E., Rose, R. (2000). Utterance verification speech recognition. IEEE Transactions on Speech and Audio Processing, vol. 8, no. 2, pp. 126–139.

Lukic, D., Jovicic, S., Punisic, S. (2011). Perception of distortion in duration of voices /š, č, ž, dž, c, r, l/. 19thTelecommunications Forum (TELFOR) Proceedings of Papers, pp. 1071 – 1074.

Lyon, R.F. (1982). A computational model of filtering, detection and compression in the cochlea, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Paris, May 1982, pp. 1282–1285.

Lyon, R.F. (1983). A computational model of binaural localization and separation, in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1983, pp. 1148–1151.

Lyon, R.F. (1984). A computational models of neural auditory processing. Proc.IEEE Int. Conf. Acoust., Speech, Signal Process., San Diego, CA, March 1984, pp. 1282-1285.

Ljolje, A., Hirschberg, J., and van Santen, J. (1996). Automatic speech segmentation for concatenative inventory selection. Progress in Speech Synthesis, Springer (1996): 305-311.

Maier, A., Hacker, C., Nöth, E., Nkenke, E., Haderlein, T., Rosanowski, F., and Schuster, M. (2006). Intelligibility of children with cleft lip and palate: Evaluation by

speech recognition techniques. 18th International Conference Pattern Recognition, I.C. Society, Ed., IEEE, pp. 477-482, 2006.

Maier, A., Honig, F., Hacker, C., Schuster, M., and Nöth, E. (2008). Automatic Evaluation of Characteristic Speech Disorders in Children with Cleft Lip and Palate. Proceedings of Interspeech, pp. 1757-1760, 2008.

Maier, A., Haderlein, T., Eysholdt, U., Rosanowski, F., Batliner, A., Schuster, M., Nöth, E. (2009). Peaks – A System for the automatic evaluation of voice and speech disorders. Speech Communication 51, 5 (May 2009), 425-437.

Manfredi, C. (2000). Adaptive noise energy estimation in pathological speech signals. IEEE Transactions on Biomedical Engineering 47(11)(2000)1538–1543.

Marple, S.L. (1987). Digital Spectral Analysis, Englewood Cliffs, NJ, Prentice-Hall, Chapter 7.

Massachusetts Eye and Ear Infirmary (1994). Voice Disorders Database,Version.1.03 [CD-ROM], Kay Elemetrics Corporation, Lincoln Park, NJ, 1994.

Massaro, D.W. (1999). Speechreading: illusion or window into pattern recognition. Trends in Cognitive Sciences, vol.3, pp.310-317, 1999.

Matassini, L., Hegger, R., Kantz, H., and Manfredi, C. (2000). Analysis of vocal disorders in a feature space. Med. Eng. Phys., vol. 22(6), pp. 413-8, 2000.

Michaelis, D., Gramss, T., Strube, H.W. (1997). Glottal-to-noise excitation ratio—a new measure for describing pathological voices. Acustica/Acta Acustica 83(1997) 700–706.

Moore, B.C.J. (1997). An Introduction to the Psychology of Hearing. London: Academic Press, 1997.

Murillo, S., Orozco, J.R., Vargas, J.F., Arias, J.D., and Castellanos, C.G. (2011). Automatic detection of hypernasality in children. Lecture Notes in Computer Science, Ed. Springer Berlin/Heidelberg, no. 6687, pp. 167-74, 2011.

Nosofsky, R.M. (1992). SIMILARITY SCALING AND COGNITIVE PROCESS MODELS. Annu. Rev. Psychol. 43:25-53.

Patterson, R.D., Holdsworth, J., and Allerhand, M. (1992). Auditory models as preprocessors for speech recognition. In The Auditory Processings of Speech: From the Auditory Periphery to Words, edited by M. E. H. Schouten (Mouton de Gruyer, Berlin), pp. 67-83, 1992.

Patterson, R.D., and Allerhand, M.H. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and software platform. Journal of Acoustical Society of America, JASA, Vol. 98, No. 4, October 1995, pp. 1890-1894, 1995.

Patterson, R.D., and Holdsworth, J. (1996). A functional model of neural activity patterns and auditory images. In: Advances in Speech, Hearing and Language Processing , (W. A. Ainsworth, ed.), Vol 3. JAI Press,London.

Patterson, R.D. (2000). Auditory images: How complex sounds are represented in the auditory system. Journal Acoustical Society of Japan, E 21, 4, 2000.

Paulo, S., Oliveira, L.C. (2003). DTW-based Phonetic Alignment Using Multiple Acoustic Features. In: Proc. Eurospeech, Geneva, Switzerland, pp. 309–312 (2003).

Paulraj, M.P, Bin Yaacob, S., Abdullah, A.N., Natraj, S.K. (2015). Segmentation of Voiced Portion for Voice Pathology Classification Using Fuzzy Logic. ResearchGate, 2015.

Pentiuc, S.G., Tobolcea, I., Schipor, O.A., Danubianu, M., & Schipor, D.M. (2010). Translation of the speech therapy programs in the Logomon assisted therapy system. Advances in Electrical and Computer Engineering, 10(2), 48-52. http://dx.doi.org/10.4316/AECE.2010.02008.

Plante, E.M., Beeson, P.M. (2007). Communication and communication disorders: A clinical introduction. Allyn & Bacon.

Punišić, S., Pantelić, S., Đoković, S., Subotić, M. (2007). Karakterizacija glasovnih odstupanja-analiza akustičkih obeležja u izgovoru frikativa /š/. Poglavlje u: Poremećaji verbalne komunikacije, prevencija, dijagnostika, tretman. M.Sovilj (ur.), IEFPG, Beograd, 62-83.S.

Punišić S., Kašić Z., Golubović, S. (2011a). Articulatory aspect of atypical voices in verbal expression. Verbal Communication Quality Interdisciplinary Research I. S. Jovicic, M. Subotic (Eds.), LAAC, IEPSP, 94-121.

Punišić S., Subotić M., Jovičić, S. (2011b). Analysis of duration and intensity of voiceless affricatives and fricatives in typical and atypical pronunciation. Third European Congress on Early Prevention, Detection and Diagnostics of Verbal Communication Disorders. Sovilj, M., Skanavis, M. i Bojanova, V. (Eds.), Proc., Belgrade, 72-79.

Punišić, S., Jovičić, S.T., Subotić, M., Šarić, Z., Bilibajkić, R. (2012). Stridens – spektralna distorzija glasova: auditivna i akustička analiza. Zbornik radova sa simpozijuma Digitalna obrada govora i slike, DOGS 2012, 4-7. oktobar 2012, Kovačica.

Punišić, S. (2012). Artikulaciono-akustički i auditivni aspekt odstupanja glasova u patološkom govoru. Doktorska disertacija, Univerzitet u Beogradu.

Qi, Y., Hillman, R.E. (1997). Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. Journal of the Acoustical Society of America 102 (1) (1997)537–543.

Öster, A.M., House, D., Protopapas, A., Hatzis, A. (2002). Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Paedia). In Proceedings from Fonetik 2002, Speech Music and Hearing Quarterly Progress and Status Report 44,45-48.

Ritchings, R.T., McGillion, M.A., Moore, C.J. (2002). Pathological voice quality assessment using artificial neural networks. Medical Engineering & Physics 24 (8) (2002)561–564.

Saenz-Lechon, N., Osma-Ruiz, V., Godino-Llorente, J.I., Blanco-Velasco, M., Cruz-Roldan, F., Arias-Londono, J.D. (2008). Effects of audio compression in automatic detection of voice pathologies, IEEE Transactions on Biomedical Engineering 55 (12)(2008)2831–2835.

Salhi, L., Talbi, M., and Cherif, A. (2008). Voice Disorders Identification Using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks. World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:2, No:9, 2008

Saz, O., Yin, S.C., Lleida, E., Rose, R., Vaquero, C., Rodrıguez, W.R. (2009). Tools and Technologies for Computer-Aided Speech and Language Therapy. Speech Communication 51 (2009) 948–967.

Schipor, O., Nestor, M. (2007). Automat parsing of audio recordings. Testing children with dyslalia. Theoretical background., Distributed Systems, "Stefan cel Mare" University of Suceava Press, Romania.

Schipor, O.A., Pentiuc, S.G., Schipor, M.D. (2012). Automatic Assessment of Pronunciation Quality of Children within Assisted Speech Therapy. Electronics and Electrical Engineering., 2012. – No. 6(122). pp. 15–18.

Schipor, O.A., Pentiuc, S.G., Schipor M.D. (2012b). Improving computer based speech therapy using a fuzzy expert system. Computing and Informatics 29 (2), 303-318, 2012.

Seneff, S. (1988). A joint synchrony/mean-rate model of auditory speech processing. Journal of Phonetics, Special Issue, Vol. 16(1), January 1988, pp. 55-76.

Shama, K., Krishna, A., Cholayya, N.U. (2007). Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. EURASIP Journal on Advances in Signal Processing 2007 (2007) 9ID85286.

Shamma, S. (1988). The acoustic features of speech sounds in a model of auditory processing: vowels and voiceless fricatives. J. Phonetics 16, 1988., 77-91.

Shao, Y., and Wang, D.L. (2008). Robust speaker identification using auditory features and computational auditory scene analysis. In Proc. IEEE ICASSP , 2008, pp. 1589–1592.

Shao, Y., Jin, Z., Wang, D., and Srinivasan, S. (2009). An auditory-based feature for robust speech recognition. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'09), 2009, pp. 4625-4628.

Shekhter, I., and Carney, L.H. (1997). A nonlinear auditory nerve model for CF-dependent shifts in tuning with sound level. Assoc. Res. Otolaryngol. 20, 617.

Shosted, R. (2006). Just put your lips together and blow? Whislted fricatives in Southern Bantu. Proc. ISSP. 565–572.

Slaney, M., and Lyon, R.F. (1993). On the importance of time – a temporal representation of sound, in Visual Representation of Speech Signals, M. Cooke, S. Beet and M. Crawford (eds.), John Wiley & Sons Ltd, 1993, pp. 95-116.

Slaney, M., Naar, D., and Lyon, R.F. (1994). Auditory model inversion for sound separation. In Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '94), vol. 2, pp. 77–80, Adelaide, Australia, April 1994.

Sollich, P., Krogh, A. (1996). Learning with ensembles: how over-fitting can be useful. In: Touretzky, D. S., Mozer, M. C., Hasselmo M. E. (Eds.), Advances in Neural Information Processing Systems 8, Denver, CO, MIT Press, Cambridge, MA, 190- 196.

Swedsen, T., Soong, F.K. (1987). On the automatic segmentation of speech signals. ICASSP-87, pp. 77-80. 1987.

Šarić, Z. (1994). Detekcija naglih promena u signalima sa AR i ARMA strukturom. Doktorska disertacija. Univerzitet u Beogradu, Elektrotehnički fakultet, 1994.

Šarić, Z., Turajlić, S. (1995). Segmentacija govora na bazi GLR ARMA mere odstojanja. Nauka Tehnika, Bezbednost, Institut Bezbednosti, Beograd, br. 1, str. 03-24, 1995.

Šarić, Z., Turajlić, S. (1995). A New ML Speech Segmentation Algorithm. Circuits Systems Signal Process, Vol. 14, No.5, pp. 615-632, 1995.

Taplidou, S.A., Hadjileontiadis, L.J. (2007). Wheeze detection based on time-frequency analysis of breath sounds. Computers in Biology and Medicine 37, 1073–1083 (2007).

Titze, I.R., Baken, R., and Herzel, H. (1993). Evidence of chaos in vocal fold vibration, New Frontiers in Basic Science, I.R. Titze. Ed. Vocal Fold Physiology, San Diego, CA: Singular Publishing Group, pp. 143-88, 1993.

Titze, I.R. (1995). Workshop on Acoustic Voice Analysis: Summary Statement, National Center for Voice and Speech, Denver, Colorado, pp. 1-36, 1995.

Toledano, D.T., Gomez, L.A., Grande, L.V. (2003). Automatic phonetic segmentation. IEEE Transactions on Speech and Audio Processing 11 (November 2003)

Turajlić, S., Šarić, Z. (1993). Novi postupak segmentacije govora na osnovu maksimalne verodostojnosti. XXXVII Jugoslovenska konf. ETAN, Komisija za automatiku (A), Beograd, 1993.

Turk, O., & Arslan, L.M. (2005). Software tools for speech therapy and voice quality monitoring. In EUSIPCO-2005. Antalya, Turkey: The 13th European Signal Processing Conference.

Valentini-Botinhao, C., Degenkolb-Weyers, S., Maier, A., Noth, E., Eysholdt, U., Bocklet, T. (2012). Automatic Detection of Sigmatism in Children . WOCCI 2012: 13-16.

Vaquero, C., Saz, O., Lleida, E., Marcos, J., Canalís, C. (2006). Vocaliza: An application for computer-aided speech therapy in Spanish language. Processings of IV Jornadas en Tecnología del Habla, Zaragoza, Spain, (2006) 321-326.

Vaquero, C., Saz, O., Lleida, E., Rodrı´guez, W.R. (2008). E-inclusion technologies for the speech handicapped. In: Proc. 2008 International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, pp. 4509–4512.

Velican, V., Strungaru, R., Grigore, O. (2012). Automatic Recognition of Improperly Pronounced Initial 'r' Consonant in Romanian. Advances in Electrical and Computer Engineering Volume 12, Number 3, 2012, 80-84.

Veprekt, P., Scordilis, M.S. (1996). A constrained DTW-based procedure for speech segmentation. In Proceedings of Sixtht Australian International Conference on Speech, Science and Technology, Adelaide, 10-12. December 1996.

Vidal, E., and Marzal, A. (1990). A review and new approaches for automatic segmentation of speech signals. In L. Torres, E. Masgrau, and M. A. Lagunas, editors, Signal Processing V: Theories and Applications, Elsevier Science Publishers B.V., 1990, pp. 43-53.

Vladisavljević, S. (1981). Poremećaj artikulacije. Privredni pregled. Beograd, 1981.

Vojnović, M., Punišić, S. (2010). Modelovanje atipičnog izgovora afrikata /c/. Zbornik radova VIII konferencije DOGS 2010, Iriški Venac, A1.6.1-A1.6.4.

Vojnović, M., Punišić, S. (2011). Atipičan izgovor frikativa /š/ kod dece. Zbornik radova, XVIII Telekomunikacioni forum TELFOR, Beograd, 1075-1078.ok

Warren, R. M. (2008). Auditory perception: An analysis and synthesis. Cambridge University Press, NY, 2008.

Wildermoth, B. R. (2001). Text-Independent Speaker Recognition Using Source Based Features, M. Phil. Thesis, Griffith University, Brisbane, Australia, 2001.

Witten, I., and Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed. San Francisco, USA: Morgan Kaufmann, 2005.

Yumoto, E., Gould, W.J., Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. Journal of the Acoustical Society of America 71(6) (1982) 1544–1550.

Zhang ,Y., and Jiang, J.J. (2004). Chaotic vibrations of a vocal fold model with a unilateral polyp. Journal of the Acoustical Society of America, vol. 115(3), pp. 1266- 9, 2004.

Zhang, Z., Jiang, J.J., Biazzo, L., and Jorgensen, M. (2005). Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. J. Voice, vol. 19(4), pp. 519-28, 2005.

Zhao, X., Shao, Y., and Wang, D.L.(2012). CASA-Based Robust Speaker Identification. IEEE Trans. Audio, Speech and Language Processing , vol.20, no.5, pp.1608-1616, 2012.

Biografija autora

Ružica Bilibajkić rođena je 15.04.1980. godine u Beogradu, gde je završila osnovnu školu i Matematičku gimnaziju. Elektrotehnički fakultet Univerziteta u Beogradu upisala je 1999. godine. Diplomirala je 2005. godine na Odseku za elektroniku, telekomunikacije i automatiku (smer telekomunikacije). Postdiplomske studije na Elektrotehničkom fakultetu u Beogradu, smer tehnička akustika, upisala je 2005. godine, gde je 2011. godine odbranila magistarsku tezu pod nazivom „Segmentacija reči na bazi MFCC i GFCC spektralnih modela“ i stekla zvanje magistra nauka.

Naučno-istraživačku karijeru, pretežno u oblasti digitalne obrade govornog signala, započela je 2006. godine kao istraživač pripravnik u Institutu za eksperimentalnu fonetiku i patologiju govora u Beogradu. Stručno i naučno interesovanje usmerila je u oblasti telekomunikacija i to na istraživanje govora i govorne komunikacije, prenosa govora, psihoakustike, produkcije i percepcije govora, patologije govora kao i mrežnih tehnologija. Trenutno je angažovana kao istraživač saradnik u Centru za unapređenje životnih aktivnosti u Beogradu. Najznačajnije rezultate postigla je u oblasti obrade govornog signala, prepoznavanja patologije govora kao i u izradi sistema za računarsku (automatsku) procenu kvaliteta govora i sluha. Do sada ima preko 25 objavljenih naučnih i stručnih radova, kao i više tehničkih i razvojnih rešenja.

Od 2006. godine učestvuje u većem broju naučnih projekata kod Ministarstva za nauku Republike Srbije. Članica je Evropskog udruženja mladih akustičara (EAA YAN - European Acoustical Association Young Acousticians Network). Nosilac je CISCO sertifikata (CSCO11489469) i to CCNA (Cisco Certified Network Asociate), CCNP Route (Cisco Certified Security Specialist) i Cisco FS (Firewall Specialist).

Prilozi

Прилог 1.

Изјава о ауторству

Потписани-а Ружица Билибајкић

број уписа /

Изјављујем

да је докторска дисертација под насловом

„Препознавање артикулационо-акустичких одступања гласова у патолошком говору "

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 29.03.2016.

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Ружица Билибајкић

Број уписа /

Студијски програм /

Наслов рада „Препознавање артикулационо-акустичких одступања гласова у патолошком говору“

Ментор др Драгана Шумарац Павловић

Потписани Ружица Билибајкић

изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 29.03.2016.

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

„Препознавање артикулационо-акустичких одступања гласова у патолошком говору “

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

(1) Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 29.03.2016.

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.