УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА У НОВОМ САДУ

РАЗВОЈ МАТЕМАТИЧКОГ МОДЕЛА ТРАЈАЊА ГЛАСОВА У АУТОМАТСКОЈ СИНТЕЗИ ГОВОРА НА СРПСКОМ ЈЕЗИКУ

ДОКТОРСКА ДИСЕРТАЦИЈА

Кандидат:
Сандра Совиљ-Никић

Ментор:
проф. др Владо Делић

Нови Сад, 2014

УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, **РБР**:	
Идентификациони број, **ИБР**:	
Тип документације, **ТД**:	Монографска документација
Тип записа, **ТЗ**:	Текстуални штампани материјал
Врста рада, **ВР**:	Докторска дисертација
Аутор, **АУ**:	Сандра Совиљ-Никић
Ментор, **МН**:	Проф. Др Владо Делић
Наслов рада, **НР**:	Развој математичког модела трајања гласова у аутоматској синтези говора на српском језику
Језик публикације, **ЈП**:	Српски (латиница)
Језик извода, **ЈИ**:	Српски/Енглески
Земља публиковања, **ЗП**:	Србија
Уже географско подручје, **УГП**:	Војводина
Година, **ГО**:	2014
Издавач, **ИЗ**:	Ауторски репринт
Место и адреса, **МА**:	Нови Сад, Трг Доситеја Обрадовића 6
Физички опис рада, **ФО**: (поглавља/страна/ цитата/табела/слика/графика/прилога)	8 поглавља /120 страна / 86 цитата / 28 табела / 74 слике
Научна област, **НО**:	Електротехника
Научна дисциплина, **НД**:	Телекомуникације и обрада сигнала
Предметна одредница/Кqучне речи, **ПО**:	Синтеза говора, моделовање трајања гласова, методе аутоматског учења
УДК	
Чува се, **ЧУ**:	Библиотека Факултета техничких наука
Важна напомена, **ВН**:	
Извод, **ИЗ**:	У оквиру ове дисертације развијено је више различитих модела трајања гласова у српском језику применом одговарајућих метода аутоматског учења. Извршена је објективна евалуација развијених модела и њихово међусобно поређење на основу квантитативних показатеља као што су RMSE(engl. root-mean-squared error), MAE (engl. mean absolute error) i CC (engl. correlation coefficient). Такође је извршено поређење модела за српски језик са перформансама модела развијених за друге језике, при чему је уочено да су перформансе модела развијених у овој дисертацији упоредљиве или чак превазилазе перформансе модела који су развијени за друге језике.
Датум прихватања теме, **ДП**:	19.06.2008.
Датум одбране, **ДО**:	

Чланови комисије, **КО**:		
Председник:	Др Олга Хаџић (редовни професор)	
Члан:	Др Драгана Бајић (редовни професор)	
Члан:	Др Слободан Јовичић (редовни професор)	Потпис ментора
Члан:	Др Маја Марковић (ванредни професор)	
Члан, ментор:	Др Владо Делић (редовни професор)	

UNIVERSITY OF NOVI SAD ● **FACULTY OF TECHNICAL SCIENCES**
21000 NOVI SAD, Trg Dositeja Obradovića 6

KEY WORDS DOCUMENTATION

Accession number, **ANO**:		
Identification number, **INO**:		
Document type, **DT**:		Monograph documentation
Type of record, **TR**:		Textual printed material
Contents code, **CC**:		PhD thesis
Author, **AU**:		Sandra Sovilj-Nikić
Mentor, **MN**:		PhD Vlado Delić
Title, **TI**:		The Development of Phone Duration Model in Speech Synthesis in the Serbian Language
Language of text, **LT**:		Serbian (Latin)
Language of abstract, **LA**:		Serbian/English
Country of publication, **CP**:		Serbia
Locality of publication, **LP**:		Vojvodina
Publication year, **PY**:		2014
Publisher, **PB**:		Author reprint
Publication place, **PP**:		Novi Sad
Physical description, **PD**: (chapters/pages/ref./tables/pictures/graphs/appendixes)		8 chapters / 120 pages / 86 references / 28 tables / 74 figures
Scientific field, **SF**:		Electrical engineering
Scientific discipline, **SD**:		Telecommunications and Signal Processing
Subject/Key words, **S/KW**:		Speech synthesis, phone duration modeling, machine learning algorithms
UC		
Holding data, **HD**:		Library of faculty of Technical Sciences
Note, **N**:		
Abstract, **AB**:		In this dissertation several different phone duration models of the Serbain language using appropriate machine learning algorithms were developed. The objective evaluation of the models obtained and their mutual comparison based on quantitative measures such as RMSE (root-mean-squared error), MAE (mean absolute error) and CC (correlation coefficient) were performed. The comparison of the models developed for the Serbian language with the performances of the models developed for other languages is also carried out. It was observed that the performances of the models developed in this dissertation are comparable or even outperform the performances of the models that have been developed for other languages.
Accepted by the Scientific Board on, **ASB**:		19.06.2008.
Defended on, **DE**:		
Defended Board, **DB**:	President:	PhD Olga Hadžić (professor)
	Member:	PhD Dragana Bajić (professor)
	Member:	PhD Slobodan Jovičić (professor)
	Member:	PhD Maja Marković (associate professor)
	Member, Mentor:	PhD Vlado Delić (professor)

Menthor's sign

SADRŽAJ

LISTA AKRONIMA

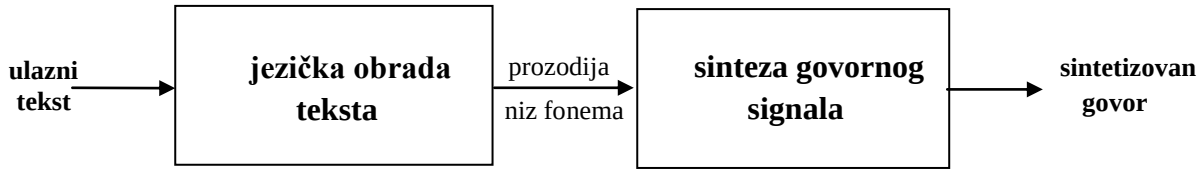
TTS	Text-to-Speech	Sinteza govora na osnovu teksta
ASR	Automatic Speech Recognition	Automatsko prepoznavanje govora
RMSE	Root-Mean-Squared-Error	Koren srednje kvadratne greške
MAE	Mean Absolute Error	Srednja apsolutna greška
CC	Correlation Coefficient	Koeficijent korelacije
CART	Classification and Regression Trees	Klasifikaciona i regresiona stabla
GTB	Gradient Tree Boosting	
SVM	Support Vector Machines	Metod potpornih vektora
VOT	Voice Onset Time	Početak vokalnih vibracija
SAMPA	Speech Assesment Method for Phonetic Alphabet	
REPTrees	Reduced Error Pruning Trees	Stabla sa potkresivanjem u cilju smanjenja greške
WEKA	Waikato Environment for Knowledge Analysis	
PDA	Personal Digital Assistant	
ARFF	Attribute-Relation File Format	
CSV	Comma Separated Values	

1. UVOD

Sinteza govora na osnovu teksta (engl. *Text-to-Speech* – TTS) je, uz automatsko prepoznavanje govora (engl. *Automatic Speech Recognition* – ASR), jedna od govornih tehnologija koja omogućava komunikaciju između čoveka i mašine putem glasa. Pomenute govorne tehnologije postaju sve aktuelnije u novije vreme, o čemu svedoči i rastuća popularnost brojnih servisa koji se upravo zahvaljujući razvoju govornih tehnologija mogu pružiti korisnicima. ASR i TTS nalaze značajnu primenu u raznim telekomunikacionim službama, kao što su čitanje *e-mail* i SMS poruka, govorni portali, iščitavanje sadržaja obimnih baza podataka, automatizacija pozivnih centara, itd. (Delić et al., 2013 a). Pored toga, govorne tehnologije nalaze svoju primenu u raznim aplikacijama zabavnog karaktera (igre na sreću, čitanje horoskopa, video igre, telefonsko glasanje, itd.), a takođe i u softveru za učenje stranih jezika. Veoma važno područje primene govornih tehnologija predstavljaju proizvodi namenjeni pomoći, informisanju i obrazovanju osoba sa posebnim potrebama, koji im omogućavaju da se osamostale i uključe u normalne tokove svakodnevnog života i na taj način zaista ostvare zakonom zagarantovana prava. Pojava govornih tehnologija omogućila je da računar čita knjige, novine sa Interneta, e-mail i SMS poruke slepim i slabovidim osobama. Takođe zahvaljujući govornim tehnologijama osobe sa oštećenjem govora su u mogućnosti da koriste telefon jer šačunar naglas čita ono što nema osoba napiše. Automatski prepoznat govor se prevodi u tekst i postaje dostupan osobama koje ne čuju što im može omogućiti da prate televizijski program. Upotreba govornih tehnologija pruža mogućnost osobama koje ne mogu da koriste ruke da govornim komandama upravljaju uređajima u okruženju. Prve primene govornih tehnologija na srpskom govornom području namenjene su slepim i slabovidim osobama i za njih je razvijeno više pomagala. AnReader je sistem za sintezu govora na srpskom jeziku koji je prvenstveno namenjen slepim i slabovidim osobama, ali mogu da ga koriste i osobe sa drugim tipovima invaliditeta. An Reader omogućava osobama sa posebnim potrebama da koriste računar. Audio biblioteka je takođe veoma koristan servis koji slepim i slabovidim osobama obezbeđuje pristup velikoj bazi knjiga preko lokalne mreže ili Interneta (Delić et al., 2013 b).

Kod TTS sistema vrši se proces konverzije teksta u govor, odnosno automatska sinteza govora na osnovu teksta, pri čemu sintetizovan govor treba da bude razumljiv i slušaocu zvuči što prirodnije. Procedura konverzije teksta u govor sastoji se iz dve osnovne faze. U prvoj fazi

vrši se jezička obrada teksta jer je ulazni tekst potrebno prevesti u oblik koji će računar direktno moći da interpretira. Druga faza podrazumeva samu sintezu govornog signala koja kao rezultat, na osnovu dobijene fonetske i prozodijske informacije, daje sintetizovan govor na izlazu. Dve pomenute faze sinteze često se u literaturi nazivaju sinteza visokog i niskog nivoa. Opšta stuktura TTS sistema prikazana je na slici 1.1.



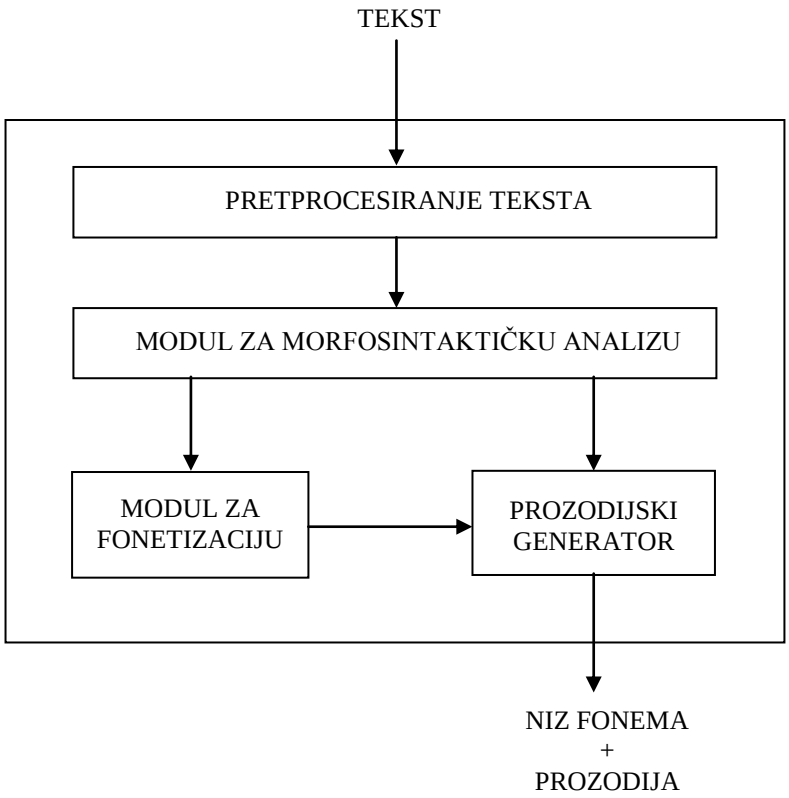
Slika 1.1 – Struktura TTS sistema

Sinteza visokog nivoa obuhvata tri osnovne faze, a to su pretprocesiranje teksta, fonetizacija i prozodijska analiza. Pretprocesiranje teksta, koji je obično u ASCII formatu, podrazumeva identifikovanje kraja rečenica i obradu znakova interpunkcije, kao i identifikaciju brojeva, skraćenica, specijalnih znakova i akronima i njihovo proširenje u ortografske reči. Nakon pretprocesiranja teksta potrebno je izvršiti njegovu fonetizaciju, s obzirom na to da sam tekst, posmatran kao niz slova, ne pruža dovoljno informacija za sintezu kvalitetnog govornog signala. Stoga, modul za jezičku obradu teksta u ovoj fazi obezbeđuje informacije o konkretnom nizu fonema koje treba izgovoriti, odnosno fonetsku transkripciju teksta. Pored fonetske transkripcije, modul za jezičku obradu teksta u okviru TTS sistema obezbeđuje i odgovarajuća prozodijska obeležja govorne celine koju treba sintetizovati. Takođe, on treba da ih u pogodnom obliku prezentuje modulu za generisanje govornog signala.

Najnovija istraživanja u oblasti sinteze govora na osnovu teksta uglavnom su usmerena ka poboljšanju njegove prirodnosti. Postizanje prirodnosti govora je od esencijalne važnosti u TTS sistemima jer prirodnost govora umnogome doprinosi i njegovoj razumljivosti. Prirodnost govora je u uskoj vezi sa njegovim prozodijskim obeležjima koja imaju ulogu u leksičkoj segmentaciji govorne celine, odnosno rastavljanju govornog niza na reči. Stoga, prirodnost govora ne predstavlja samo stvar estetike, već doprinosi i razumevanju poruke upućene slušaocu.

Pored uloge u leksičkoj segmentaciji prozodijska obeležja imaju još jednu funkciju na leksičkom nivou, ona doprinose ispravnoj intonaciji rečenice. U srpskom jeziku, koji spada u grupu jezika koji nemaju fiksan akcenat, značaj ispravne intonacije je još izraženiji. S obzirom na to da položaj i/ili tip akcenta u okviru reči ponekad utiče na značenje same reči ili označava

razliku u vrednosti morfološke kategorije iste leksičke reči, pogrešno akcentovana reč može otežati percepciju ili potpuno blokirati komunikaciju, ukoliko je produkovana reč zamenjena drugom leksičkom jedinicom.



Slika 1.2 – Modul za jezičku obradu teksta

Imajući u vidu složenu akcenatsku strukturu srpskog jezika, kao i relativno slobodnu akcentuaciju, kvalitetnu sintezu govora na osnovu teksta moguće je ostvariti jedino pomoću obimnog akcenatskog rečnika (Sečujski, 2002). Međutim, zbog pomenutih slučajeva višeznačne akcentuacije, pored akcenatskog rečnika neophodan je i modul za morfosintaktičku analizu rečenice, koji bi bio u stanju da razreši sve dileme koje se javljaju u slučaju da je neku reč u rečenici moguće akcentovati na više različitih načina. Takođe, modul za morfosintaktičku analizu rečenice bi, s jedne strane, davao neophodne informacije modulu za fonetizaciju, a s druge strane modulu za generisanje akustičke reprezentacije prozodije (slika 1.2).

Sa stanovišta govornih tehnologija i njihove primene među najznačajnija prozodijska obeležja ubrajaju se osnovna učestanost, glasnost i trajanje govornih segmenata. Iako su mnoga do sada sprovedena istraživanja neosporno pokazala da kretanje osnovne učestanosti glasa, odnosno f_0 kriva, predstavlja najznačajnije prozodijsko obeležje govorne celine sa perceptivnog

stanovišta, takođe postoje istraživanja koja pokazuju da trajanja govornih segmenata imaju neznatno manje značajnu ulogu od f_0 krive za razumevanje poruke upućene slušaocu. Štaviše, precizno određena f_0 kriva i prirodno trajanje fonema zajedno mnogo više doprinose poboljšanju kvaliteta sintetizovanog govora nego što to čini sama f_0 kriva (Bulyko et al., 1999).

Trajanje glasova u prirodnom govoru u velikoj meri zavisi od konteksta u kojem se dati govorni segment nalazi, pri čemu je ta zavisnost izuzetno kompleksna i uključuje mnogobrojne faktore (Klatt, 1976).

Proučavanje vremenske organizacije govora i uticaja različitih fonoloških, sintaksnih, fizioloških i drugih faktora na trajanje govornih segmenata je od izuzetne važnosti kako za razumevanje procesa proizvodnje govora, tako i za razvoj sinteze govora u cilju proizvodnje sintetizovanog govora visokog kvaliteta (van Santen, 1992). Stoga je veoma bitna komponenta TTS sistema specijalizovani modul (engl. *duration system*) čiji je zadatak modelovanje trajanja govornih segmenata iz prirodnog govora uzimajući u obzir različite faktore. Identifikovanje najuticajnijih faktora predstavlja krucijalan korak u procesu modelovanja trajanja jer izbor neadekvatnog ili nepotpunog skupa faktora može dovesti do značajne greške predikcije trajanja. Imajući u vidu prirodu problema, skup faktora koji opisuju kontekst u kome se određeni segment nalazi sačinjavaju samo oni faktori koje je moguće izvući iz samog teksta. Značaj i način uticaja određenog faktora na trajanje direktno zavise od konkretnog jezika, te se skup najuticajnijih faktora razlikuje od jezika do jezika. Izbor ovog skupa zahteva određeni stepen lingvističkog znanja, kao i određenu statističku analizu govornog korpusa u cilju određivanja što tačnijih vrednosti trajanja glasova za dati jezik, imajući u vidu značaj trajanja glasova sa perceptivnog stanovišta (Klatt, 1987;van Santen, 1993).

S obzirom na značaj trajanja govornih segmenata sa perceptivnog stanovišta, odnosno na ulogu trajanja u razumevanju izgovorenog teksta, specijalizovani modul za određivanje potrebnog trajanja predstavlja komponentu TTS sistema od izuzetne važnosti za proizvodnju sintetizovanog govora visokog kvaliteta. Modelovanje trajanja govornih segmenata u različitim jezicima jeste predmet mnogobrojnih do sada sprovedenih istraživanja u kojima su primenjivane različite tehnike modelovanja. Modeli za predviđanje trajanja mogu se podeliti u dve osnovne grupe: modeli zasnovani na primeni pravila (engl. *rule-based models*) i korpusno orijentisani modeli (engl. *corpus-based models*).

Jedan od najpoznatijih modela za predviđanje trajanja primenom niza uzastopnih pravila, ujedno i najstariji model, razvio je Denis Klat (Denis Klatt). Njegov rad predstavlja osnovu za razvoj modela trajanja za nekoliko svetskih jezika kao što su američki engleski, švedski,

francuski i brazilski portugalski. Osnovna prednost ovakvih modela leži u činjenici da ne zahtevaju obiman govorni korpus, što je bilo od izuzetne važnosti u vreme njihovog nastanka, kada računarski resursi za generisanje i analizu obimnih govornih korpusa nisu bili dostupni kao danas. Međutim, razvojem računarske tehnologije, korpusno orijentisani modeli postaju sve zastupljeniji. Korpusno orijentisani statistički modeli zahtevaju obiman korpus snimljenog govora jer se modelovanje trajanja vrši primenom neke od metoda automatskog učenja na obimnom govornom korpusu.

Razvoj matematičkog modela trajanja glasova u automatskoj sintezi govora na srpskom jeziku primenom metoda automatskog učenja na osnovu obimnog korpusa snimljenog govora, kao i identifikovanje najuticajnijih faktora na trajanje glasova u srpskom jeziku predstavljaju predmet istraživanja u okviru ove doktorske disertacije.

1.1 CILJ ISTRAŽIVANJA I SADRŽAJ DOKTORSKE DISERTACIJE

Primarni cilj istraživanja u okviru ove doktorske disertacije jeste dobiti matematički model trajanja glasova za automatsku procenu potrebnog trajanja glasova u sintezi govora na srpskom jeziku primenom metoda automatskog učenja a uzimajući u obzir faktore koji u najvećoj meri utiču na trajanje. S obzirom na to da izbor neadekvatnog ili nepotpunog skupa atributa može rezultovati velikom greškom prilikom procene trajanja, identifikovanje najuticajnijih faktora predstavlja krucijalan korak u procesu modelovanja trajanja glasova, tako da je još jedan od ciljeva ove disertacije bio dobijanje skupa najuticajnijih faktora.

U prvom poglavlju dat je kratak opis sinteze govora i istaknut značaj prozodijskih obeležja za razumevanje poruke upućene slušaocu. Takođe je ukazano na značaj trajanja glasova kao jednog od prozodijskih obeležja, kao i na važnost modelovanja trajanja govornih segmenata iz prirodnog govora, uzimajući u obzir najuticajnije faktore, u cilju proizvodnje sintetizovanog govora visokog kvaliteta.

Trajanje govornih segmenata, koje predstavlja veoma bitno prozodijsko obeležje sa aspekta govornih tehnologija i njihove primene, kao i faktori koji ga određuju opisani su u drugom poglavlju.

Treće poglavlje daje pregled najznačajnijih dosadašnjih istraživanja u oblasti modelovanja trajanja govornih segmenata ukazujući na prednosti i mane različitih tipova modela.

U četvrtom poglavlju data je klasifikacija fonema u srpskom jeziku i objašnjena je karakteristika zvučnosti fonema. Prikazane su artikulaciono-akustičke karakteristike vokala i konsonanata, kao i značaj akustičkih karakteristika za percepciju sintetizovanog glasa. U ovom poglavlju takođe je opisan akcenat srpskog standardnog jezika i percepcija akcenata. S obzirom na to da su prilikom modelovanja trajanja glasova uzeti u obzir silabički faktori koji deluju na nivou sloga i javljaju se kao posledica organizovanja fonema u slogove u ovom poglavlju predstavljen je i algoritam za podelu reči na slogove u srpskom jeziku. Pomenuti algoritam razvijen je za potrebe istraživanja u okviru ove doktorske disertacije.

Peto poglavlje posvećeno je opisu govorne baze korišćene u procesu modelovanja trajanja glasova. Takođe je dat detaljan opis skupa faktora koji u najvećoj meri utiču na trajanje govornih segmenata u srpskom jeziku a koji su korišćeni prilikom razvoja modela trajanja glasova.

Metode automatskog učenja koje su u okviru ove disertacije korišćene u procesu modelovanja trajanja glasova u srpskom jeziku ukratko su prikazane u šestom poglavlju.

U sedmom poglavlju ukratko je opisan softverski paket WEKA (Hall et al., 2009) koji je korišćen prilikom razvoja modela trajanja glasova u srpskom jeziku. U ovom poglavlju dat je prikaz i analiza dobijenih rezultata, odnosno objektivna evaluacija razvijenih modela i njihovo međusobno poređenje na osnovu kvantitativnih pokazatelja kao što su RMSE (engl. *root-mean-squared-error*), MAE (engl. *mean absolute error*) i CC (engl. *correlation coefficient*). Takođe je dato poređenje dobijenih rezultata sa rezultatima koji se odnose na druge jezike i prethodno razvijene modele.

Zaključak i pravci daljeg istraživanja navedeni su u osmom poglavlju.

2. TRAJANJE FONEMA I FAKTORI KOJI GA ODREĐUJU

Govor nastaje kao proizvod funkcionisanja govornih organa, što predstavlja njegovu fiziološku osnovu koju čine pokreti i položaji govornih organa prilikom obrazovanja glasova u akustičkom domenu. On je stoga po svojoj prirodi artikulaciono-akustička pojava. Uzimajući u obzir akustička, fiziološka i funkcionalna svojstva glasa kao osnovne jedinice ljudskog govora, glas se može definisati kao zvuk koji predstavlja proizvod govornih organa i ima sposobnost da diferencira značenja (Šipka, 2008). Kao fonem definiše se onaj glas čijom zamenom u datoj reči ona menja lingvističko značenje. Dakle, fonem je glas koji ima diferencijalno-semantičku funkciju. Drugim rečima, svaki zvuk koji je proizvod govornih organa i čuje se kao glas, a nema takvu funkciju, nije fonem.

U lingvističkom smislu, fonem predstavlja osnovnu jedinicu u govornoj komunikaciji za dati jezik. Redosled fonema u nizu određen je pravilima jezika, a njihov broj razlikuje se od jezika do jezika. Prosečan broj fonema nalazi se u opsegu između 20 i 37 (de Boer, 2000). U fonetskom sistemu srpskog jezika postoji 30 fonema koji su u ćiriličnom pismu reprezentovani sa 30 slova a u latiničnom sa 27 slova i 3 digrama. Proučavanjem glasova kao govornih jedinica za obeležavanje razlike u značenju reči bavi se fonologija. Fonetika se kao posebna nauka u okviru opšte lingvistike bavi izučavanjem fonema kako u domenu anatomije (artikulatorna fonetika), tako i u domenu fizike (akustička fonetika). Takođe postoji i treća grana fonetike, auditivna fonetika, koja se bavi recepcijom i razumevanjem govornih nizova.

Sa aspekta govornih tehnologija i njihove primene veoma bitno prozodijsko obeležje jeste trajanje govornih segmenata. Trajanje govornog segmenta predstavlja određeni vremenski interval, pri čemu se pod trajanjem ne podrazumevaju samo početni i završni trenutak govornog segmenta, već i trenuci svih relevantnih događaja u okviru njih, koji zavise od tipa govornog segmenta koji predstavlja osnovnu govornu jedinicu. Primera radi, ukoliko je u sintezi govora osnovna govorna jedinica dvoglas, tada je bitno i u kom trenutku se dešava prelaz iz jednog glasa u drugi.

Svaki fonem poseduje određena artikulaciono-akustička svojstva koja zavise od načina i mesta artikulacije fonema i kao takva razlikuju jedan fonem od drugog. Fonemi smešteni u različite kontekste uvek sa sobom nose i ispoljavaju svoje artikulaciono-akustičke karakteristike. Trajanje fonema, koje predstavlja veoma bitno prozodijsko obeležje, između ostalog zavisi i od načina i mesta artikulacije fonema. Ukoliko se dva različita fonema nađu u istom fonetskom

kontekstu, kreirajući tako minimalni par, oni će imati različito trajanje koje je upravo posledica različitih artikulaciono-akustičkih osobina tih fonema. Ovako definisano, tzv. *inherentno trajanje* fonema, ne može se smatrati apsolutnom veličinom jer svako konkretno merenje uključuje i niz drugih faktora koji takođe utiču na konkretno izmereno trajanje govornog segmenta. Stoga, inherentno trajanje se može tretirati kao faktor koji utiče na konkretno izmereno trajanje govornog segmenta i treba ga razlikovati od prosečnog trajanja izmerenog u nekim konkretnim uslovima. Izmereno prosečno trajanje fonema bliže je inherentnom trajanju ukoliko je broj izvršenih merenja trajanja određenog fonema dovoljno velik tako da je uzet u obzir i uticaj ostalih faktora kao što su okolni fonemi, tempo artikulacije, naglašenost sloga itd.

Inherentno trajanje vokala predstavlja ono što izaziva razliku u trajanju različitih vokala u istom fonetskom kontekstu. Ono je u direktnoj korelaciji sa artikulacionim karakteristikama vokala, odnosno sa položajem i oblikom jezika, otvorom vilice i oblikom usana. Inherentno trajanje konsonanata, kao i inherentno trajanje vokala, predstavlja ono što uslovljava razliku u trajanju različitih konsonanata koji se nalaze u istom fonetskom okruženju.

Trajanje govornih segmenata ograničeno je sa dva kraja, odnosno oni ne mogu trajati proizvoljno kratko niti se mogu produžavati preko neke mere. U oba slučaja, razlog ograničenja trajanja je fiziološke prirode. Prilikom produkcije određenog govornog segmenta artikulacijski pokreti imaju određenu brzinu koja je uslovljena nizom fizioloških osobina. Stoga, minimalno trajanje segmenta ne može biti kraće od vremena potrebnog za realizovanje pokreta. Takođe, ograničenje trajanja govornog segmenta pored produkcije uslovljeno je i percepcijom govora, jer je za percepciju bilo koje akustičke dimenzije potrebno odgovarajuće minimalno vreme. Na drugom kraju, trajanje daha ograničava produžavanje govornih segmenata.

2.1 LINGVISTIČKI FAKTORI

Budući da se govor realizuje u vremenu, opis vremenske dimenzije govora predstavlja važan deo njegovog akustičkog opisa. Istraživanje vremenske organizacije govora (engl. *speech timing*) postaje intenzivnije u poslednjim decenijama prošlog veka, što je svakako posledica razvoja sinteze govora i potrebe za automatskom procenom prirodnog trajanja glasova u cilju postizanja što veće prirodnosti sintetizovanog govora.

Na trajanje govornih segmenata utiču mnogobrojni lingvistički i paralingvistički faktori. Prema nivou delovanja, lingvistički faktori mogu se svrstati u tri grupe: segmentni, silabički i suprasilabički (White, 2002).

Segmentni faktori su oni faktori koji izazivaju razliku u trajanju dva različita fonema koji postavljeni u isti kontekst kreiraju minimalni par. Različito inherentno trajanje je upravo posledica različitih artikulacionih karakteristika fonema, tj. uslovljeno je različitim položajem i oblikom jezika, otvorom vilice i oblikom usana kod vokala, odnosno načinom i mestom artikulacije konsonanata. Klatt identifikuje razliku u trajanju dugih i kratkih vokala u engleskom jeziku, kao i duže trajanje bezvučnih frikativa od zvučnih. On je u svojim istraživanjima takođe utvrdio da bilabijalni plozivi traju duže od alveolarnih i velarnih ploziva (Klatt, 1976) Takođe, primećena je i povezanost položaja jezika sa inherentnim trajanjem vokala u srpskom (Sovilj-Nikić, 2007) , hrvatskom (Bakran, 1996) i slovenačkom jeziku (Gros, 2000). Naime, zatvoreniji vokali, kod kojih je jezik bliže nepcu, traju kraće od otvorenijih vokala koji su udaljeniji od konsonantske artikulacije. Vokal /a/ koji je u srpskom jeziku najotvoreniji, odnosno vokal koji je prema visini jezika nizak vokal, ima najduže trajanje. S druge strane, vokal /i/ koji je prema visini jezika u usnoj duplji visok vokal ima najkraće trajanje (Sovilj-Nikić, 2007). Način artikulacije konsonanata takođe uslovljava razlike u njihovom trajanju, pri čemu unutar iste kategorije (klase) postoje razlike u trajanju zavisno od mesta artikulacije i zvučnosti fonema. U hrvatskom jeziku plozivi duže traju od nazala, bezvučni plozivi traju duže od zvučnih, labijali traju duže od alveolara i dentala (Bakran, 1996). Sve navedene distinkcije uslovljene različitom artikulacijom fonema jesu deo i predmet proučavanja fonetike određenog jezika.

Silabički faktori deluju na nivou sloga i javljaju se kao posledica organizovanja fonema u slogove. Klatt identifikuje produženje trajanja vokala i konsonanata u naglašenim slogovima kao jedan od perceptivno najistaknutijih silabičkih faktora u engleskom jeziku (Klatt, 1976). Duže trajanje akcentovanih vokala u odnosu na njihove neakcentovane parnjake primećeno je takođe u srpskom (Sovilj-Nikić, 2007; Marković & Bjelaković, 2011; Lehiste & Ivić, 1996), hrvatskom (Bakran, 1996) i slovenačkom jeziku (Gros, 2000). Razlika u trajanju naglašenih i nenaglašenih slogova je lingvističke prirode i predstavlja posledicu delovanja fonoloških pravila u jeziku. Klatt napominje da je razlika u trajanju između naglašenih i nenaglašenih slogova najveća na kraju fraze, što ukazuje na interakciju između silabičkog i suprasilabičkog nivoa. Broj i tip okolnih segmenata, u okviru sloga i između slogova, takođe utiču na trajanje fonema. Konsonanti u grupama (klasterima) kraće traju nego konsonanti u intervokalskom položaju što je još izraženije u naglašenim slogovima (White, 2002; Bakran, 1996). Uticaj zvučnosti konsonanta na trajanje vokala ispred njega utvrđen je u mnogim jezicima i smatra se univerzalnim principom (Sovilj-Nikić, 2007; Bakran, 1996; Gros, 2000). Jedno od tumačenja ove pojave jeste da vokali traju obrnuto proporcionalno energiji potrebnoj za artikulaciju narednog konsonanta što je u skladu sa teorijom o konstantnoj energiji artikulacije sloga (Chen, 1970). Klatt navodi da je

produženje trajanja vokala ukoliko mu sledi zvučni konsonant izraženije u slogovima ispred sintaktičke granice, kao i da trajanje vokala na toj poziciji može imati perceptivnu ulogu u razlikovanju zvučnih i bezvučnih konsonanata kode (Klatt, 1976).

Uticaj lingvističke strukture i posledicu organizovanja slogova u reči i konstituente hijerarhijski viših nivoa moguće je posmatrati kao suprasilabičke faktore. Klatt je identifikovao veći broj ovakvih faktora u engleskom jeziku i njihov uticaj klasifikovao u tri grupe: 1) produženje u blizini granice, 2) produženje usled isticanja i 3) skraćenje zbog veličine konstituenta. Efekat produženja na kraju fraze Klatt smatra najvažnijim zbog njegovog lingvističkog značaja, kao i perceptivnog isticanja. Produženje trajanja vokala ispred sintaktičke granice, odnosno vokala i konsonanata ukoliko je finalni slog zatvoren primećeno je u srpskom (Sovilj-Nikić, 2007; Lehiste & Ivić, 1996), hrvatskom (Bakran, 1996), francuskom, engleskom (Oller, 1973), kao i u mnogim drugim jezicima i takođe se smatra univerzalnom pojavom. U srpskom jeziku produženje je najveće na kraju nefinalne klauze, zatim na kraju upitne rečenice, a najmanje je na kraju izjavne rečenice (Lehiste & Ivić, 1996). Veličina produženja takođe zavisi i od tipa sloga. Najveće je produženje vokala u poslednjem otvorenom slogu. Međutim, budući da je produženjem obuhvaćen čitav slog, vokal u poslednjem zatvorenom slogu ispred pauze duži je od ostalih vokala u nenaglašenim slogovima.

Klatt kao drugi veoma značajan efekat suprasilabičkih faktora navodi produženje trajanja usled rečeničnog naglašavanja. U srpskom jeziku je utvrđeno da je posledica isticanja neke reči u rečenici njeno duže trajanje, pri čemu je u izjavnim rečenicama obično inicijalna reč u fokusu ukoliko je u pitanju subjekat u rečenici (Lehiste & Ivić, 1996).

Veličina naglasne celine (stope), koju određuje broj slogova u stopi, takođe predstavlja jedan od faktora koji utiču na trajanje vokala u srpskom jeziku jer primećeno je da vokali srpskog jezika, bez obzira da li su naglašeni ili ne, traju kraće ukoliko je broj slogova u stopi veći (Sovilj-Nikić, 2007; Marković & Milićev, 2009).

U grupu graničnih efekata Klatt još ubraja produženje na početku i na kraju reči, produženje na kraju izjave, kojim je obuhvaćeno nekoliko slogova, kao i produženje trajanja na kraju pasusa, koje se odnosi na poslednju rečenicu u pasusu. Produženje trajanja vokala i konsonanata na početku i na kraju reči primećeno je i u holandskom jeziku (Klabbers, 2000).

Smatra se da produženje trajanja ne samo vokala, već i konsonanata u poslednjem slogu u reči, kada reči slede u nizu i kada se ne može primetiti nikakva pauza na granici reči ima perceptivnu funkciju, odnosno da doprinosi označavanju granica između reči (Bakran, 1996). Stoga, ovaj efekat treba uzeti u obzir u sintezi govora jer on ne doprinosi samo prirodnosti govora, nego i njegovoj razumljivosti.

Uticaj različitih paralingvističkih faktora kao što su brzina izgovaranja i način i stil govora obično se zanemaruju prilikom modelovanja trajanja pod pretpostavkom da se mogu smatrati nepromenljivim u okviru čitave govorne baze (van Santen, 1994).

2.2 FAKTORI U RAZLIČITIM JEZICIMA

Kao što je ranije napomenuto, trajanje fonema u prirodnom govoru zavisi od mnogobrojnih faktora, pri čemu je uticaj ovih faktora u različitim jezicima različit. Stoga, imajući u vidu direktnu zavisnost najuticajnijih faktora na trajanje govornih segmenata od konkretnog jezika, izbor skupa atributa razlikuje se od jezika do jezika. Na slici 2.1 prikazan je broj obeležja koja su autori koristili prilikom modelovanja trajanja govornih segmenata u različitim jezicima. U tabeli 2. 1 navedeni su najuticajniji faktori koji su korišćeni u procesu modelovanja trajanja govornih segmenata u engleskom (Campbell, 1992), nemačkom (Moebius & van Santen, 1996), japanskom (Venditti & van Santen 1998), katalonskom (Febrer et al., 1998), češkom (Batušek, 2002), grčkom (Lazaridis et al., 2007), turskom (Öztürk, 2005), litvanskom (Norkevičius & Raškinis, 2008) i hindu jeziku (Krishna & Murthy, 2004).

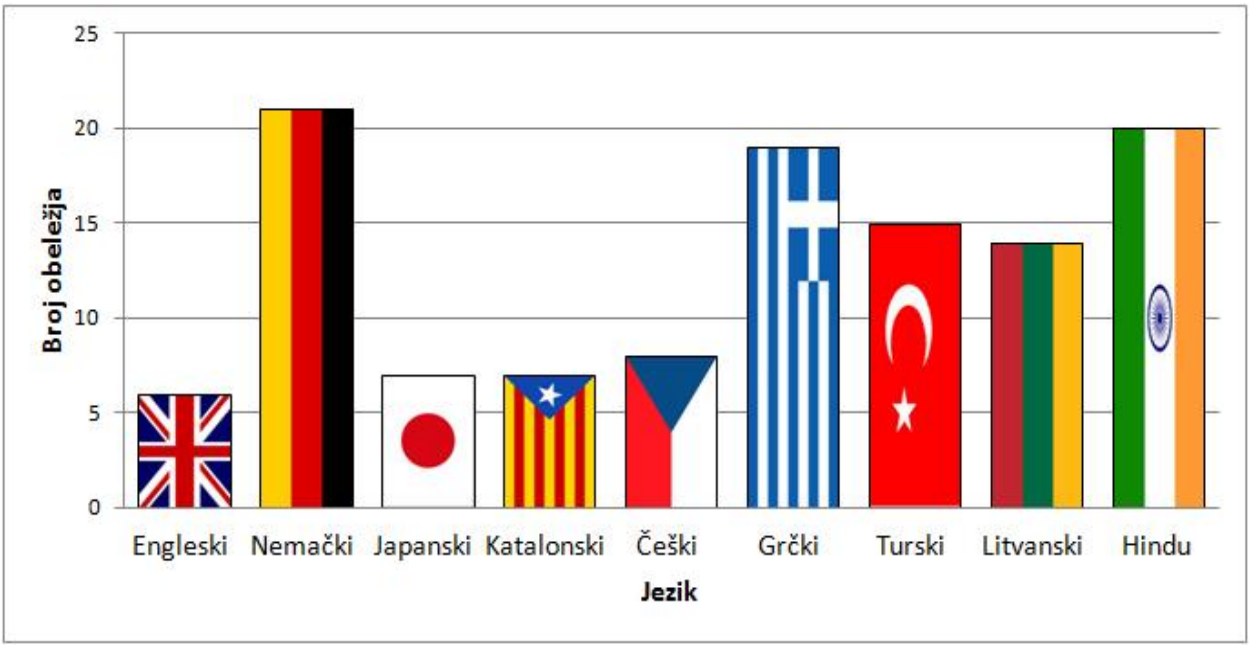


Tabela 2.1 – Najuticajniji faktori u nekoliko različitih jezika

JEZIK	NAJUTICAJNIJI FAKTORI
ENGLESKI	broj fonema u slogu, tip nosioca sloga, položaj sloga u stopi, položaj sloga u frazi, leksički akcenat, klasa reči (funkcijska reč, leksička reč)
NEMAČKI	identitet fonema, tip govornog segmenta (vokali: prednji/srednji/zadnji, konsonanti: zvučni/bezvučni, frikativi, nazali, laterali, itd.), klasa reči (funkcijska reč, konstituentska reč, složenica), položaj fraze u izjavi, dužina fraze (broj reči), položaj reči u frazi (početni/srednji/krajnji), dužina reči (broj slogova), položaj sloga u reči (početni/srednji/krajnji), leksički akcenat (primarni, sekundarni, nenaglašen), položaj segmenta u slogu (onset, nukleus, koda), fonetski kontekst (identitet prvog, drugog i trećeg fonema sa leve/desne strane datog fonema), tip fonetskog konteksta (tip prvog, drugog i trećeg fonema sa leve/desne strane datog fonema)
JAPANSKI	identitet fonema, identitet prethodnog/narednog fonema, levi prozodijski kontekst (da li je slog na početku određene fraze ili ne), desni prozodijski kontekst (da li je slog na kraju određene fraze ili ne), rečenični naglasak, struktura sloga (otvoren/zatvoren)
KATALONSKI	identitet fonema, leksički akcenat, položaj sloga u frazi, fonetsko okruženje (prethodni/naredni fonem), dužina sloga, položaj sloga
ČEŠKI	identitet fonema, fonetsko okruženje (prethodni/naredni fonem), broj fonema u slogu/reči/frazi, položaj fonema u slogu (od početka/kraja), položaj fonema u reči (od početka/kraja), položaj reči u frazi (početni/srednji/krajnji)
GRČKI	identitet fonema, položaj sloga u reči (početni/srednji/krajnji/jednosložna reč), broj fonema pre vokala u datom slogu (veličina onseta), broj fonema nakon vokala u datom slogu (veličina kode), broj slogova u reči, položaj fonema u slogu (od početka/kraja), da li je fonem prvi/poslednji u slogu, redni broj sloga u reči, da li je fonem pre ili posle vokala u slogu (onset/koda), mesto artikulacije fonema, da li je slog naglašen, nivo pauze nakon sloga, leksički akcenat sloga, broj slogova od prethodne pauze, broj slogova do naredne pauze, broj naglašenih slogova od prethodne pauze, broj naglašenih slogova do naredne pauze, vrsta reči, klasa reči (funkcijska reč, leksička reč)
TURSKI	identitet fonema, leksički akcenat, položaj fonema u slogu (onset, nukleus, koda), tip sloga, nivo pauze, položaj sloga u reči, položaj reči/sloga u rečenici, dužina reči (broj slogova), ukupan broj reči (slogova) u rečenici (reči), ukupan broj reči od (do) prethodne (naredne) pauze, ukupan broj slogova od (do) prethodne (naredne) pauze
LITVANSKI	identitet fonema, identitet dva prethodna/naredna fonema, dužina sloga, položaj fonema u slogu (od početka/kraja), dužina reči, položaj sloga u reči (od početka/kraja), dužina fraze, položaj reči u frazi (od početka/kraja), da li je fonem prvi/poslednji u reči
HINDU	identitet fonema, obeležja fonema (npr. visina vokala, tip konsonanta, zvučnost konsonanta, itd.), obeležja prethodnog fonema - levi kontekst, obeležja narednog fonema - desni kontekst, struktura sloga, položaj u slogu, da li je fonem prvi/poslednji u slogu, položaj sloga u reči (početni/srednji/krajnji/jednosložna reč), broj slogova u reči, redni broj sloga u reči (prvi, drugi, itd.), nivo pauze nakon sloga (slog je unutar reči, na kraju reči/fraze/rečenice), dužina fraze (broj reči), položaj fraze u izjavi, broj fraza u izjavi

Tabela 2.1 – Najuticajniji faktori u nekoliko različitih jezika

3. MODELOVANJE TRAJANJA GLASOVA

U prirodnom govoru trajanja glasova su izuzetno zavisna od konteksta u kojem se određeni govorni segment nalazi, pri čemu je ta zavisnost veoma kompleksna i uključuje mnogobrojne faktore (van Santen, 1992). Stoga je za proizvodnju sintetizovanog govora visokog kvaliteta veoma bitno da u okviru TTS sistema postoji specijalizovani modul čiji je zadatak da modeluje trajanja govornih segmenata iz prirodnog govora uzimajući u obzir različite faktore koji utiču na trajanje, a imajući u vidu prirodu problema, to su oni faktori koje je moguće izvući iz samog teksta. Modelovanje trajanja govornih segmenata u različitim jezicima jeste predmet mnogobrojnih do sada sprovedenih istraživanja u kojima su primenjivane različite tehnike modelovanja. Modeli za predviđanje trajanja mogu se podeliti u dve grupe: modeli zasnovani na primeni pravila (engl. *rule-based models*) i korpusno orijentisani modeli (engl. *corpus-based models*).

3.1 MODELI ZASNOVANI NA PRIMENI PRAVILA

Jedan od najpoznatijih modela za predviđanje trajanja primenom niza uzastopnih pravila, ujedno i najstariji model, razvio je Denis Klatt (Klatt, 1976). Kod ovakvog tipa modela pretpostavlja se da svaki fonem poseduje određeno inherentno trajanje koje predstavlja jedno od distinktivnih svojstava fonema. Svakom fonemu inicijalno se dodeljuje inherentno trajanje koje se zatim modifikuje primenom niza uzastopnih pravila. Primenom određenog pravila trajanje datog govornog segmenta produžava se ili skraćuje za određeni procenat, pri čemu nakon skraćenja trajanje ne može biti manje od minimalnog trajanja.

Trajanje fonema prema Klattovom modelu određuje se kao:

$$d = d_{\min} + r \cdot (d_{inh} - d_{\min}) \tag{3.1}$$

gde je: $d_{\min}$ minimalno trajanje datog fonema (ukoliko je naglašen)

d_{inh} inherentno trajanje datog fonema

r faktor korekcije

U slučaju primene N pravila faktor korekcije iznosi:

$$r = \prod_{i=1}^{N} r_i \tag{3.2}$$

pri čemu primena svakog pojedinačnog pravila unosi korekciju r_i .

Na osnovu uočenih perceptivno značajnih pravilnosti, prikazanih u tabeli 3.1 (Klatt, 1976), koje se odnose na trajanje govornih segmenata Klatt predlaže skup pravila čijom primenom modeluje trajanje fonema u britanskom engleskom. Kao što je već ranije napomenuto primena određenog pravila doprinosi smanjenju ili povećanju trajanja govornog segmenta za određeni procenat. Kod ovakvog tipa modela neophodna je *lookup* tabela u kojoj se nalaze minimalno i inherentno trajanje svakog fonema.

1. Produženje trajanja vokala i konsonanata koji se nalaze u slogu ispred pauze
2. Skraćenje trajanja svih segmenata (uključujući vokale i konsonante) u slogu koji se ne nalazi ispred pauze
3. Skraćenje trajanja segmenata u slogu ukoliko taj slog nije poslednji slog u reči
4. Konsonanti koji se ne nalaze na početku reči se skraćuju
5. Nenaglašeni fonemi i fonemi sa sekundarnim akcentom se skraćuju
6. Naglašeni vokali se produžavaju
7. U zavisnosti od fonetskih osobina konteksta u kome se nalaze trajanje vokala može biti skraćeno ili produženo
8. Skraćenje trajanja konsonanata u klasterima
9. Skraćenje trajanja vokala ukoliko im sledi bezvučni konsonant

Tabela 3.1 – Različiti perceptivno značajni efekti

Klattov rad predstavlja osnovu za razvoj modela trajanja govornih segmenata za nekoliko svetskih jezika kao što su američki engleski (Allen et al., 1987), švedski (Carlson & Granstrom, 1986), francuski (Bartkova & Sorin, 1987), italijanski (Boula de Mareüil, 1999), nemački (Kohler, 1988), brazilski portugalski (Simoes, 1990), grčki (Epitropakis et al., 1993).

Prilikom razvoja modela zasnovanih na primeni pravila neophodno je znanje stručnih lingvista za dati jezik, odnosno njihovo učestvovanje u sastavljanju određenih pravila. Pisanje

pravila često može biti veoma naporan i vremenski zahtevan posao, a takođe veoma je teško formirati dovoljan broj pravila kojima bi bile obuhvaćene sve moguće situacije u nekom jeziku. Stoga, kod ovakvog tipa modela pojava izuzetaka najčešće predstavlja problem jer su pravila uglavnom takva da često dovode do prevelikog uopštavanja. Takođe, nedostatak ovakvih modela je i što obično ne uzimaju u obzir korelaciju između faktora koji utiču na trajanje govornih segmenata a koja je veoma često prisutna (Rao & Yegnanarayana, 2007). Međutim, pored niza prethodno spomenutih nedostataka modela zasnovanih na primeni pravila oni poseduju i određene prednosti. Osnovna prednost ovakvih modela leži u činjenici da ne zahtevaju obiman govorni korpus, što je bilo od izuzetne važnosti u vreme njihovog nastanka kada računarski resursi za generisanje i analizu obimnih govornih korpusa nisu bili dostupni kao danas.

3.2 KORPUSNO ORIJENTISANI MODELI

Razvojem računarske tehnologije korpusno orijentisani modeli postaju sve zastupljeniji. Korpusno orijentisani statistički modeli zahtevaju obiman korpus snimljenog govora, jer se modelovanje trajanja vrši primenom neke od metoda automatskog učenja na obimnom govornom korpusu. U zavisnosti od metode automatskog učenja koja se primenjuje u svrhu modelovanja trajanja van Santen (van Santen, 1993) razlikuje tri tipa modela:

- linearni statistički modeli
- modeli dobijeni primenom neuralnih mreža
- modeli dobijeni primenom stabala odluke (engl. *decision trees*).

Van Santen, kao primer linearnog statističkog modela, navodi aditivni model koji je Kaiki razvio za japanski jezik. Kod ovakvog modela trajanje govornog segmenta u datom kontekstu dobija se sumiranjem niza parametara kojima se modeluje uticaj različitih kontekstualnih faktora (identitet fonema, fonetsko okruženje, tj. prethodni i naredni fonem, itd.) na trajanje govornog segmenta. Estimacija ovih parametara vrši se primenom neke od standardnih statističkih metoda (Kaiki et al., 1990). Takođe, postoje i multiplikativni modeli kod kojih se umesto sumiranja parametara vrši njihovo množenje.

U svojim radovima van Santen predlaže model sume proizvoda koji predstavlja generalizaciju aditivnog i multiplikativnog modela (van Santen, 1995). Ovaj model zasnovan je na linearnoj statistici i pretpostavci invarijantnosti smera delovanja određenog faktora, odnosno na pretpostavci da je smer promene trajanja usled delovanja određenog faktora uvek isti bez obzira na uticaj drugih faktora. Takođe, uzima se u obzir i činjenica da je uticaj određenih

faktora na različite grupe fonema različit. Prema modelu sume proizvoda, trajanje fonema koji je u određenom kontekstu predstavljen preko odgovarajućeg vektora obeležja d određuje se kao:

$$DUR(d) = \sum_{i \in T} \prod_{j \in I_i} S_{i,j}(d_j) \tag{3.3}$$

pri čemu T predstavlja skup indeksa odgovarajućeg proizvodnog činioca, dok I_i predstavlja skup indeksa faktora koji se pojavljuju u *i*-tom činiocu proizvoda. Parametri $S_{i,j}$ nazivaju se skalirajući faktori i predstavljaju uticaj odgovarajućih faktora i i j, a d_j je *j*-ti elemenat vektora obeležja d. Kod aditivnog modela $T = \{1,...,n\}$ i $I_i = \{1\}$ a kod multiplikativnog modela $T = \{1\}$ i $I_1 = \{1,...,n\}$.

Modelovanje trajanja primenom modela sume proizvoda obično podrazumeva tri uzastopna koraka [37]:

1. formiranje kategorijskog stabla, odnosno podela prostora obeležja u odgovarajuće kategorije
2. za svaki list (završni čvor) stabla razvija se odgovarajući model sume proizvoda
3. estimacija parametara modela

Van Santen kao osnovne prednosti predloženog modela ističe mogućnost opisivanja međusobnog uticaja faktora jednostavnim aritmetičkim operacijama sabiranja i množenja, kao i sposobnost modela za estimaciju parametara u slučaju pojave retkih vektora. Prednost ovakvih modela je što ne zahtevaju obiman govorni korpus, dok je njihov nedostatak odsustvo automatizacije, odnosno potreba za nadgledanjem celokupnog procesa razvoja modela od strane eksperta. Ovakav model primenjen je za modelovanje trajanja govornih segmenata u engleskom (van Santen, 1995), nemačkom (Moebius & van Santen, 1996), japanskom (Venditti & van Santen, 1998), holandskom (Klabbers, 2000), francuskom, italijanskom, španskom, mandarinskom (Weibin et al., 2000).

Druga mogućnost za obučavanje sistema u cilju sticanja određenih znanja koja će se kasnije koristiti za predviđanje trajanja govornih segmenata jeste primena neuralnih mreža. U svom radu Campbell prvi primenjuje neuralne mreže za predviđanje trajanja sloga u engleskom jeziku i predlaže modelovanje u dva koraka (Campbell, 1992). U prvom koraku on primenjuje neuralne mreže sa tri nivoa i propagacijom unazad (engl. *three-level back-propagation neural networks*) za predviđanje odstupanja trajanja pojedinačnog sloga od srednje vrednosti. U cilju pronalaženja faktora koji utiču na trajanje sloga Campbell primenjuje analizu kategorijskog faktora. Trajanje svakog segmenta u slogu određuje se u drugom koraku modelovanja rešavanjem sledeće jednačine po k:

$$\Delta = \sum_{i=1}^{n} \exp(\mu_i + k \cdot \sigma_i) \tag{3.4}$$

gde je: Δ trajanje sloga određeno u prethodnom koraku

n broj segmenata u slogu

μ_i i σ_i srednja vrednost i standardna devijacija trajanja datog fonema

Nakon primene iterativnog postupka za rešavanje eksponencijalne jednačine (3.4) odgovarajućem govornom segmentu i dodeljuje se trajanje $\exp(\mu_i + k \cdot \sigma_i)$.

Ovakav postupak matematičkog mapiranja između vektora obeležja i trajanja sloga veoma je sličan regresionoj analizi, ali takođe omogućava i modelovanje nelinearnosti. Proces razvoja modela trajanja govornih segmenata primenom neuralnih mreža praktično je potpuno automatizovan, što je još jedna od njegovih prednosti. S druge strane, nedostatak primene ovakvog postupka je u njegovoj neobjašnjivosti, odnosno unošenje bilo kakvih promena zahteva ponovnu dugotrajnu obuku sistema. Takođe, ovakav postupak modelovanja zahteva obimnu govornu bazu i veoma je osetljiv na pojavu retkih vektora obeležja.

Riedi je takođe koristio neuralne mreže za modelovanje trajanja govornih segmenata u nemačkom jeziku, dok je Cordoba primenio neuralne mreže za predviđanje trajanja fonema španskog jezika (Öztürk, 2005). Neuralne mreže takođe su primenjene i za modelovanje trajanja fonema u finskom jeziku (Vainio, 2001).

U treću grupu statističkih modela spadaju modeli zasnovani na primeni stabala odluke. Prvi takav model za predviđanje trajanja govornih segmenata u američkom engleskom jeziku razvio je Riley (Riley, 1992) koristeći CART (engl. *Classification and Regression Trees*) tehniku (Breiman et al., 1984). Modelovanje primenom CART tehnike predstavlja specijalan slučaj modelovanja zasnovanog na primeni stabala. Ovakav tip statističkog modelovanja podrazumeva sukcesivnu podelu prostora obeležja u cilju minimizacije greške predikcije. Primenom CART tehnike formira se binarno stablo koje u svakom čvoru sadrži da/ne pitanje o nekom obeležju, odnosno faktoru koji utiče na trajanje govornog segmenta. Počevši od korena stabla u svakom koraku u fazi obuke formiraju se sva moguća pitanja za svako od mogućih obeležja. U svakom čvoru algoritam statistički selektuje najznačajnije obeležje kao i pitanje vezano za to obeležje, odnosno za dato obeležje bira se ono pitanje koje podatke deli tako da je nakon izvršene podele sličnost među podacima u novonastalom čvoru najveća. Opisani postupak rekurzivno se ponavlja nad svakim podskupom sve dok ne bude zadovoljen unapred postavljeni kriterijum za prestanak podele unutar date particije. Predviđanje trajanja govornog segmenta vrši se prolaskom kroz

stablo odluke, od korena do lista stabla, prolazeći kroz unutrašnje čvorove stabla onom putanjom koja se formira u zavisnosti od zadovoljenja određenog uslova o vrednostima obeležja u svakom od unutrašnjih čvorova. List stabla sadrži predviđenu vrednost trajanja datog govornog segmenta.

Riley je prilikom modelovanja trajanja govornih segmenata u američkom engleskom koristio ručno labeliranu govornu bazu koja sadrži 1500 rečenica, a kao najznačajnija obeležja on izdvaja identitet fonema, fonetsko okruženje, akcenat, položaj u reči i položaj u frazi usvajajući kategorizaciju fonema prema načinu i mestu artikulacije u cilju smanjenja broja uticajnih faktora.

CART tehnika razvijena kao spoj statistike i veštačke inteligencije poseduje niz prednosti i kao takva danas predstavlja jednu od najčešće primenjivanih metoda za modelovanje trajanja govornih segmenata. Jedna od osnovnih prednosti CART algoritma jeste mogućnost validacije razvijenog modela, što se u praksi najčešće vrši procenom performansi modela na podacima koji nisu korišćeni u fazi obuke. Takođe, CART algoritam je relativno robustan u slučaju manjka podataka (Breiman et al., 1984), omogućava jednostavnu interpretaciju i obradu dobijenih rezultata, statistički selektuje najznačajnija obeležja i omogućava kombinovanje kategorijskih (npr. identitet fonema) i numeričkih vrednosti (npr. trajanje fonema) obeležja. Postupak modelovanja trajanja primenom CART metode takođe zahteva obiman korpus snimljenog govora kao i druge korpusno orijentisane metode modelovanja trajanja. Jedan od nedostataka kod ovakvog tipa modelovanja trajanja jeste što se obično dobija veliko stablo $T_{\max}$ koje može biti formirano striktno prema podacima koji su korišćeni u fazi obuke i takvo stablo nema sposobnost generalizacije, odnosno neće pokazati dobre performanse u slučaju primene nad podacima koji nisu korišćeni u fazi obuke. Stoga, potrebno je pronaći stablo optimalne veličine i izbeći *overfitting* podataka. U literaturi se navodi da je bilo niz pokušaja za prevazilaženje ovog problema među kojima se kao najbolje rešenje izdvaja Breimanov postupak (Breiman et al., 1984) koji se sastoji od nekoliko koraka:

1) formira se sekvenca podstabala $T_{\max} \supseteq \ldots \supseteq T_k \supseteq \ldots \supseteq T_K = t_1$

2) za svako podstablo procenjuje se stopa greške

3) bira se stablo sa najmanjom stopom greške, odnosno stablo optimalne veličine

Opisani postupak naziva se potkresivanje stabla (engl. *cost-complexity pruning*). Prilikom formiranja sekvence podstabala koja se dobijaju odstranjivanjem pojedinih grana parametar

kompleksnosti α varira od 0 (za $T_{\max}$) do ∞ (za podstablo koje sadrži samo koren) tako da je zadovoljen uslov:

$$\min_{T} \left[\sigma^2(T) + \alpha \cdot |T|\right] \tag{3.5}$$

gde je: $\sigma^2(T)$ varijansa greške predikcije za dato podstablo

$|T|$ broj terminalnih čvorova podstabla

U cilju procene stope greške podstabla se testiraju na podacima koji nisu korišćeni u fazi obuke. Procedura koja se najčešće primenjuje za procenu naziva se ukrštena validacija (engl. *cross-validation*). Naime, ukupna količina raspoloživih podataka podeli se na deset međusobno disjunktnih podskupova na kojima se vrši testiranje podstabala koja su formirana na osnovu preostalih 9/10 podataka. S obzirom da se postupak testiranja ponavlja deset puta za svako podstablo se izračunava prosečna varijansa. Ukoliko se varijansa dobijena na ovaj način posmatra kao funkcija veličine stabla, tada će za stablo određene veličine biti dostignut minimum varijanse i takvo stablo smatra se stablom optimalne veličine, jer dalje povećanje veličine stabla povećava varijansu.

Modelovanje trajanja govornih segmenata primenom CART metode realizovano je za mnoge jezike, među koje spadaju češki (Batušek, 2002), grčki (Lazaridis et al., 2007), turski (Öztürk, 2005), litvanski (Norkevičius & Raškinis, 2008), mandarinski (Yu, 2005), britanski engleski (Bouzon & Hirst, 2002), srpskohrvatski (Sečujski et al., 2011), vijetnamski (Mixdorff et al., 2005), korejski (Lee & Oh, 1999; Chung, 2002; Chung & Huckvale, 2001), indijski jezici hindu i telugu (Krishna & Murthy, 2004; Krishna et al., 2004), finski (Vainio, 2001).

U novijim istraživanjima, pored regresionih stabala, prilikom razvoja modela trajanja govornih segmenata sve više se primenjuju i druge metode automatskog učenja (engl. *machine learning*). Bajesove mreže (engl. *Bayesian networks*) primenjene su za predikciju trajanja fonema u engleskom jeziku i dobijeni su dobri rezultati čak i u slučaju kada nedostaju vrednosti nekih obeležja (Goubanova & King, 2008). Prilikom modelovanja trajanja fonema u grčkom jeziku primenjeni su algoritmi zasnovani na uzorcima (engl. *instance-based*) (Lazaridis et al., 2007). Kod ovakvog tipa modela u fazi predikcije koristi se funkcija odstojanja u cilju pronalaženja člana skupa za obuku, koji je najbliži segmentu čije se trajanje u datom trenutku određuje. GTB (engl. *Gradient Tree Boosting*) metod primenjen je prilikom modelovanja trajanja fonema u japanskom, mandarinskom i engleskom jeziku (Yamagishi et al., 2008) kao alternativa konvencionalnom pristupu modelovanja primenom regresionih stabala. GTB algoritam je meta algoritam koji se bazira na formiranju višestrukih regresionih stabala i

korišćenju prednosti takvog pristupa. Prilikom modelovanja trajanja sloga u indijskom jeziku korišćen je SVM (engl. *Support Vector Machines*) metod (Rao & Yegnanarayana, 2005). Kod modelovanja trajanja fonema u grčkom jeziku primenjen je takođe SVM regresioni metod, kao i mnoge druge metode automatskog učenja među koje spadaju linearna regresija, modelska stabla, regresiona stabla, meta algoritmi kao što su aditivna regresija i *bagging* (Lazaridis et al., 2010). U cilju poboljšanja tačnosti predikcije trajanja fonema u grčkom jeziku Lazaridis (Lazaridis et al., 2011) u svom radu predlaže regresionu fuzionu tehniku baziranu na kombinaciji predikcija različitih individualnih modela. Statistički model koji predstavlja kombinaciju regresionih stabala i linearnih regresionih modela primenjen je za modelovanje trajanja govornih segmenata u japanskom jeziku (Iwahashi & Sagisaka, 2000). Ovakav pristup bazira se na pretpostavci da se različiti algoritmi ponašaju različito u različitim uslovima, te stoga nedostaci svojstveni jednom algoritmu bivaju nadomešćeni prednostima drugog.

Postojeći sintetizator govora na srpskom jeziku, razvijen u okviru projekta AlfaNum na Fakultetu tehničkih nauka u Novom Sadu (Sečujski et al., 2007) iako trenutno najkvalitetniji TTS sistem na srpskom jeziku, zahvaljujući modularnosti i otvorenosti sistema za dalju doradu, može biti i jeste predmet stalnih istraživanja koja se sprovode u cilju njegovog unapređenja, odnosno poboljšanja prirodnosti sintetizovanog govora. U poslednjih nekoliko godina vršena su istraživanja u okviru kojih je razvijen modul za predikciju trajanja fonema u srpsko-hrvatskom jeziku baziran na primeni regresionih stabala (Sečujski et al., 2011). Detaljnije poređenje pomenutog modela sa modelima razvijenim u okviru ove doktorske disertacije za predviđanje trajanja glasova u srpskom jeziku biće dato u nekom od narednih poglavlja.

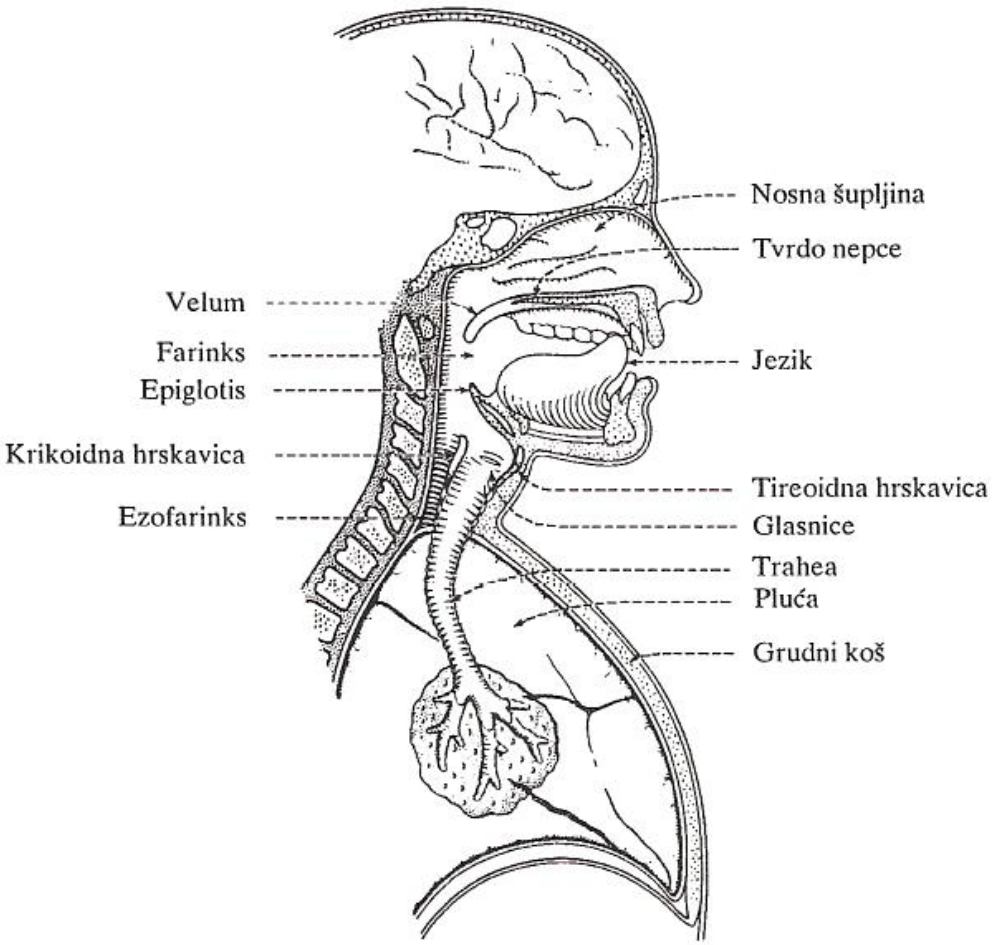
4. SRPSKI JEZIK I NJEGOVE SPECIFIČNOSTI

Srpski jezik koji danas govori nešto više od 15 miliona ljudi u Srbiji i širom sveta spada u grupu indoevropskih, južnoslovenskih jezika. Srpski jezik jeste jedan od vidova srpskohrvatskog jezika koji zavisno od sredine u kojoj se govori ima različite vidove i nazive (Stanojčić et al., 2005). Hrvatski jezik i bošnjački jezik, kojima se danas govori u Hrvatskoj, odnosno Bosni i Herzegovini, jesu takođe vidovi srpskohrvatskog jezika. Srpskohrvatski jezik čine tri narečja: štokavsko, čakavsko i kajkavsko. Narečja su nazive dobila po upitnoj zamenici za stvari, koja u najvećem delu srpskohrvatskog jezika glasi *što*, a u druga dva znatno manja dela – *ča* i *kaj*. Štokavsko narečje zauzima najveći deo srpskohrvatske jezičke teritorije. Ono se prostire u celoj Srbiji, Crnoj Gori, Bosni i Herzegovini i u velikom delu Hrvatske. Prema tome kojim je glasovima zamenjen stari glas "jat", štokavsko narečje se deli na *ekavski, (i)jekavski i ikavski izgovor*, a prema razvitku akcenta i oblika – na *starije štokavske dijalekte i mlađe štokavske dijalekte (novoštokavske)*. *Šumadijsko-vojvođanski dijalekat***,** jedan od novoštokavskih dijalekata, koji je zastupljen u najvećem delu severozapadne Srbije, u Sremu, Šumadiji, najvećem delu Bačke i Banata postao je osnova srpskog standardnog jezika ekavskog izgovora (Stanojčić et al., 2005).

4.1 KLASIFIKACIJA FONEMA I KARAKTERISTIKA ZVUČNOSTI

U formiranju govora učestvuju vokalni i nazalni trakt i pluća sa bronhijama i trahejom. Vokalni trakt sačinjavaju glasne žice, ždrelo (veza jednjaka i usta), usna duplja i usne. Pluća, bronhije i traheja predstavljaju izvor energije za stvaranje govora. Šematski prikaz govornog mehanizma dat je na slici 4.1 (Flanagan, 1972). Govorni signal predstavlja akustički talas koji se izrači iz sistema kada se strujanje vazduha izbačenog iz pluća uobliči raznim suženjima u vokalnom traktu. Promenom oblika vokalnog trakta u vremenu dolazi do formiranja različitih glasova. Kao fonem definiše se onaj glas čijom zamenom u datoj reči ona menja lingvističko značenje. U lingvističkom smislu, fonem predstavlja osnovnu jedinicu u govornoj komunikaciji za dati jezik. Proučavanjem glasova kao govornih jedinica za obeležavanje razlike u značenju reči bavi se fonologija. Fonetika se kao posebna nauka u okviru opšte lingvistike bavi izučavanjem fonema kako u domenu anatomije (artikulatorna fonetika), tako i u domenu fizike (akustička fonetika). Artikulatorna fonetika analizira i opisuje sve anatomske detalje koji

neposredno učestvuju u generisanju jednog fonema. S druge strane, akustička fonetika analizira foneme u akustičkom domenu. Ona opisuje svaki fonem skupom akustičkih obeležja koja su u direktnoj korelaciji sa pozicijom artikulacionih organa, kao i načinom artikulacije. U fonetskom sistemu srpskog jezika postoji 30 fonema koji su u ćiriličnom pismu reprezentovani sa 30 slova a u latiničnom sa 27 slova i 3 digrama. Svaki od 30 fonema, koji postoje u fonetskom sistemu našeg jezika, poseduje određeni skup karakteristika koji omogućava jedinstvenu identifikaciju fonema. Međutim, određena karakteristika može biti svojstvena većem broju fonema, što omogućava klasifikaciju fonema u određene grupe i to na više načina. Uobičajena klasifikacija fonema jeste prema zvučnosti, mestu artikulacije i načinu artikulacije. Analiza percepcije fonema najčešće se vrši na bazi ove tri karakteristike koje se u literaturi obično nazivaju osnovne distinktivne karakteristike (Jovičić, 1999).



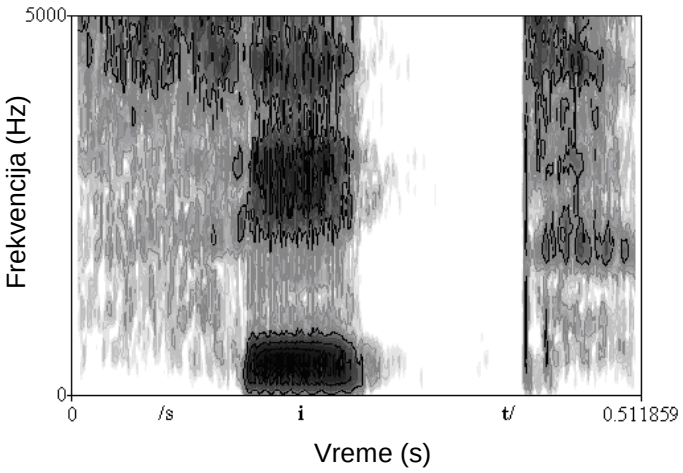
Slika 4.1 – Šematski prikaz govornog mehanizma

Zvučnost fonema podrazumeva aktivnost glasnih žica u toku generisanja fonema. Prema toj klasifikaciji fonemi se dele na zvučne i bezvučne. U srpskom jeziku ima dvadeset zvučnih i deset bezvučnih fonema koji se klasifikuju kao što je prikazano u tabeli 4.1 (Jovičić, 1999).

<table>
<tr><th colspan="3">ZVUČNI</th><th>BEZVUČNI</th></tr>
<tr><th colspan="2">formantna struktura</th><th>laringealna zvučnost</th><td rowspan="3">/p/, /t/, /k/,
/s/, /š/, /f/, /h/,
/c/, /ć/, /č/</td></tr>
<tr><th>vokali</th><th>sonanti</th><td rowspan="2">/b/, /d/, /g/,
/v/, /z/, /ž/,
/đ/, /dž/</td></tr>
<tr><td>/i/, /e/,
/a/, /o/, /u/</td><td>/m/, /n/, /nj/,
/r/, /l/, /lj/, /j/</td></tr>
<tr><td>VOKALI</td><td>SONANTI</td><td colspan="2">KONSONANTI</td></tr>
</table>

Tabela 4.1 – Klasifikacija fonema

U tabeli 4.1 data je detaljnija podela zvučnih fonema izvršena na osnovu spektralne karakteristike zvučnosti, koja, u akustičkom pogledu, zavisi od oblika vokalnog trakta. Kod vokala /a/, /e/, /i/, /o/ i /u/ u vokalnom traktu se formiraju veliki akustički rezonatori. Rezonantne učestanosti akustičkog filtra (vokalnog trakta) nazivaju se *formanti* i u spektru se ističu kao pojačana spektralna područja (slika 4.2). U spektru sonanata /m/, /n/, /nj/, /r/, /l/, /lj/ i /j/ takođe postoji izražena formantna struktura, ali su viši formanti znatno slabijeg intenziteta. Za razliku od vokala, sonanti imaju i šumni deo spektra. Preostalih osam zvučnih fonema /b/, /d/, /g/, /v/, /z/, /ž/, /đ/ i /dž/ poseduju tzv. laringealnu zvučnost. Naime, kod njih je zvučnost izražena spektralnim koncentratom neposredno iznad osnovne laringealne učestanosti. Međutim, ovaj spektralni koncentrat ne može se smatrati formantom. Kod fonema /v/ ovaj spektralni koncentrat je nešto širi nego kod ostalih, te se u određenim okolnostima može smatrati da fonem /v/ ima poluformantnu strukturu spektra i zbog toga se on ponekad svrstava u sonante (Jovičić, 1999).



*Slika 4.2 – Spektrogram reči **sit** u kojoj je vokal /i/ pod kratkim akcentom*

Na osnovu toga da li poseduju akustičku osobinu zvučnosti ili ne, u srpskom standardnom jeziku postoji sistem parova zvučnih i bezvučnih suglasnika (tabela 4.2). Naime, samo zvučnost prvog fonema u skupu fonema *bod* razlikuje taj skup od skupa *pod*, tj. reč *bod* od reči *pod*. Karakteristika zvučnosti nije apsolutno stabilna karakteristika fonema. U našem jeziku, u jednoj reči, ne može ispred bezvučnog suglasnika biti izgovoren zvučni suglasnik i obrnuto. U takvim situacijama primenjuje se pravilo jednačenja suglasnika po zvučnosti, odnosno zvučni suglasnik se zamenjuje njegovim bezvučnim parnjakom i obrnuto, kako u promeni oblika (*poljubac – poljupca*), tako i u tvorbi reči (*pretpostaviti (pred + postaviti*)).

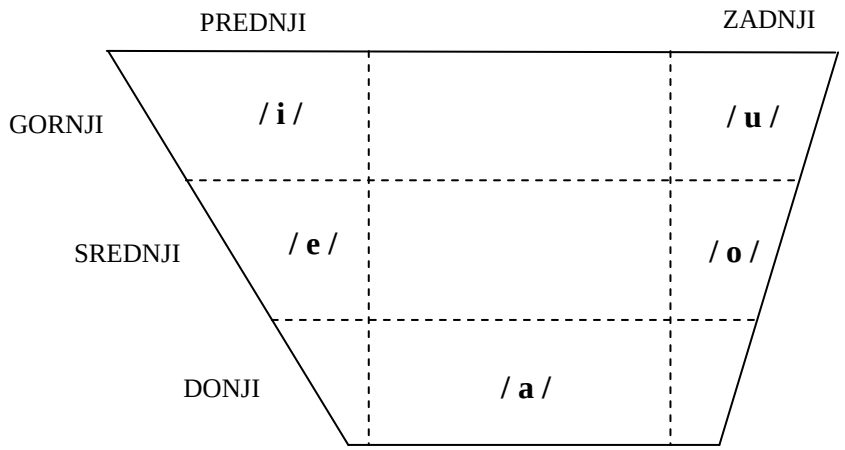
ZVUČNI SUGLASNICI	/b/	/g/	/d/	/đ/	/ž/	/z/	/dž/	/v/		
BEZVUČNI SUGLASNICI	/p/	/k/	/t/	/ć/	/š/	/s/	/č/	/f/	/h/	/c/

Tabela 4.2 – Sistem parova suglasnika u srpskom književnom jeziku stvoren na osnovu opozicije akustičkih osobina zvučnost-bezvučnost

Prema položaju i ulozi govornih organa pri artikulaciji svi glasovi se mogu podeliti u dve grupe. U prvu grupu spadaju samoglasnici (vokali), a u drugu suglasnici (konsonanti). Samoglasnici (vokali) /i/, /e/, /a/, /o/ i /u/ imaju akustičku osobinu vokalnosti (lat. *vocalis* – zvučan, glasan) zasnovanu na slobodnom prolasku fonacione struje kroz usnu duplju i njenom delovanju na sluh. Pošto su prilikom artikulacije samoglasnika govorni organi otvoreni i nigde se ne dodiruju, odnosno nema mesta dodira (lat. *locus* – mesto), samoglasnici su nelokalizovani. S druge strane, suglasnici su lokalizovani glasovi, jer se pri njihovoj artikulaciji govorni organi na nekom mestu dodiruju stvarajući tako prepreku fonacionoj struji. Pravi suglasnici (konsonanti) /b/, /p/, /g/, /k/, /d/, /t/, /đ/, /ć/, /ž/, /š/, /z/, /s/, /dž/, /č/, /v/, /f/, /h/ i /c/ imaju akustičku osobinu konsonantnosti (lat. *consonans* – suglasnik) – šumnosti. Ovu osobinu karakteriše postojanje prepreke fonacionoj struji, čime se za sluh stvara jak utisak šuma. U treću grupu glasova spadaju glasovi koji su sa aspekta prethodno navedenih dvaju akustičkih osobina neutralni, odnosno niti imaju osobinu vokalnosti, niti osobinu konsonantnosti. Takvi glasovi nazivaju se glasnici (sonanti). Kod artikulacije sonanata /m/, /n/, /nj/, /r/, /l/, /lj/, /j/ fonaciona struja slobodno prolazi pored prepreka koje postoje.

4.2 SAMOGLASNICI (VOKALI)

Samoglasnici su glasovi pri čijem izgovoru vazduh izbačen iz pluća prolazi kroz dušnik i u grkljanu pokreće glasne žice, zatim slobodno teče kroz usnu duplju i izlazi u prostor oko govornika. S obzirom na to da vazdušna struja ne nailazi na prepreke, vokali se čuju kao čisti, zvučni tonovi. U srpskom standardnom jeziku postoji pet vokala: /i/, /e/, /a/, /o/ i /u/. Na slici 4.3 (Jovičić, 1999) šematski je prikazan položaj jezika u usnoj duplji pri generisanju vokala. Položaji jezika u horizontalnoj dimenziji na slici su označeni kao prednji/zadnji, a u vertikalnoj dimenziji kao gornji/srednji/donji.



Slika 4.3 – Šematski prikaz položaja jezika u usnoj duplji pri generisanju vokala

4.2.1 Artikulacione karakteristike vokala

Fonetski raspored vokala direktno zavisi od položaja govornih organa i on je u našem jeziku sledeći: /i/, /e/, /a/, /o/ i /u/. U generisanju vokala učestvuju glasne žice, meko (zadnje) nepce, pokreti jezika, pokreti donje vilice i usne. Za vokale je, pored oblika usana, karakteristična i aktivnost celog tela jezika unutar vokalnog trakta.

Pri izgovoru vokala /i/ vrh jezika naslonjen je na donje sekutiće i jezik se pokreće u pravcu prednjeg dela usne duplje izdižući se visoko prema tvrdom (prednjem) nepcu. Zato se /i/ naziva visokim vokalom prednjeg reda. Usne su razvučene i blago rastavljene. Razmak između vilica pri izgovoru vokala /i/ je vrlo mali, odnosno vilični ugao je najoštriji. Stoga, prema veličini protoka vazdušne struje /i/ je zatvoreni vokal.

Kod generisanja vokala /e/ jezik se takođe kreće prema prednjem delu usne duplje, ali se izdiže do srednje visine. Stoga, vokal /e/ se naziva srednjim vokalom prednjeg reda. Usne su pri

izgovoru /e/ otvorene, ali skoro nepokretne. Vokal /e/ je srednje otvorenosti jer je razmak između gornje i donje vilice srednji, odnosno vilični ugao je nešto veći nego pri izgovoru vokala /i/.

Pri izgovoru vokala /a/ jezik je u skoro horizontalnom položaju i malo povučen unazad u odnosu na vokal /e/. Vokal /a/ je niski vokal srednjeg reda jer jezik ostaje nisko na dnu usne duplje. Usne su pri izgovoru vokala /a/ otvorene i skoro nepokretne. Prema veličini ugla koji čine gornja i donja vilica /a/ je najotvoreniji vokal, odnosno pri njegovom izgovoru usta su najviše otvorena.

Pri izgovoru vokala /o/ jezik se kreće prema zadnjem delu usne duplje. Vokal /o/ je srednji vokal zadnjeg reda jer se jezik, povlačeći se unazad, izdiže do srednje visine prema nepcima. S obzirom na veličinu viličnog ugla, odnosno otvorenost usta vokal /o/ je srednje otvoren. Pri izgovoru vokala /o/ usne igraju aktivnu ulogu i dobijaju elipsast oblik.

Vokal /u/ je takođe vokal zadnjeg reda jer se pri njegovom izgovoru jezik kreće ka zadnjem delu usne duplje. Za razliku od vokala /o/ kod izgovora vokala /u/ jezik se izdiže visoko prema nepcima, te je stoga /u/ visoki vokal zadnjeg reda. Prema veličini ugla koji zauzimaju gornja i donja vilica, odnosno prema otvorenosti usta /u/ je zatvoren vokal. Usne se pri izgovoru /u/ isturaju napred i zaokružuju.

Opisane artikulacione karakteristike vokala pokazuju da prema položaju jezika postoje dva prednja (/i/ i /e/), jedan srednji (/a/) i dva zadnja (/o/ i /u/) vokala (slika 4.3). Prema veličini protoka vazdušne struje postoje dva zatvorena (/i/ i /u/), dva srednje otvorena (/e/ i /o/) i jedan otvoren (/a/) vokal, a prema visini jezika u usnoj duplji postoje dva visoka (/i/ i /u/), dva srednja (/e/ i /o/) i jedan nizak (/a/) vokal.

4.2.2 Akustičke karakteristike vokala

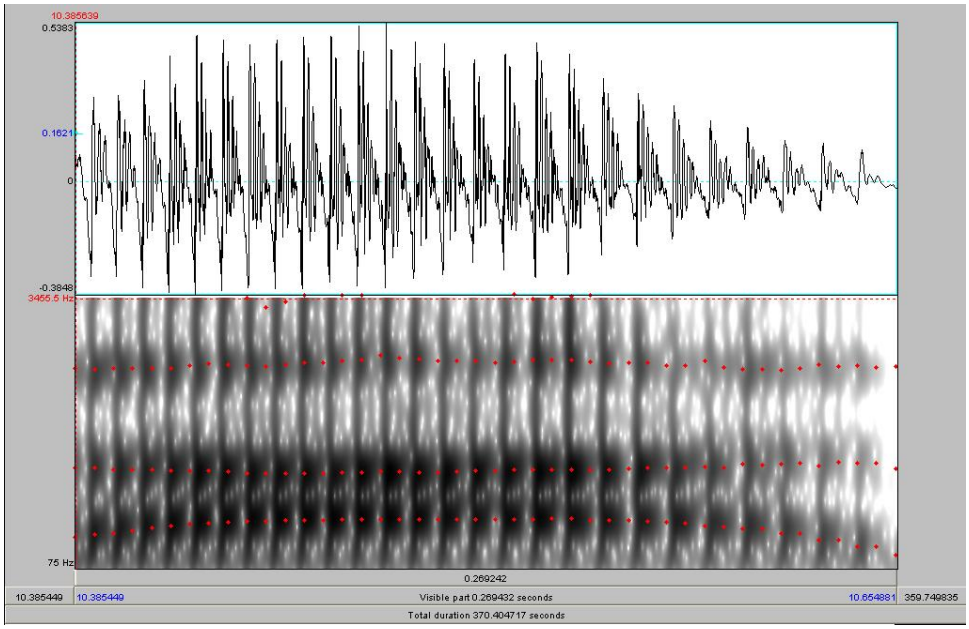
Kao što je ranije napomenuto, proučavanjem fonema u akustičkom domenu bavi se posebna grana fonetike – akustička fonetika. Akustičke osobine glasova, tj. one fizičke osobine glasova koje prima i prepoznaje slušalac svojim čulom sluha kao akustički utisak, usko su povezane sa artikulacijom glasova. Akustička fonetika opisuje svaki fonem skupom akustičkih obeležja koja su u direktnoj korelaciji sa pozicijom artikulacionih organa i kao takva daje teorijsku i eksperimentalnu osnovu za obradu govornog signala i njenu primenu u telekomunikacijama, prepoznavanju i sintezi govora, u dijagnostici patologije govora itd.

U zavisnosti od toga kako se fonaciona struja kreće kroz govorni aparat i kakav utisak stvara u sluhu onoga koji je percipira, vokali srpskog standardnog jezika imaju pored već

spomenute osobine *vokalnosti* i neke dodatne akustičke osobine (Stanojčić et al., 2005). U nastavku su navedene akustičke osobine kao i vokali koji ih poseduju:

1. akustička osobina kompaktnosti (lat. *compactus* – zbijen): /a/;
2. akustička osobina difuznosti (lat. *difundere* – raspršiti, raširiti): /i/, /u/;
3. akustička osobina gravisnosti (lat. *gravis* – dubok, taman): /o/, /u/;
4. akustička osobina akutnosti (lat. *acutus* – oštar): /i/, /e/;
5. akustička osobina neprekidnosti: /i/, /e/, /a/, /o/, /u/;
6. akustička osobina zvučnosti: /i/, /e/, /a/, /o/, /u/;
7. akustička osobina napregnutosti: /i/, /u/.

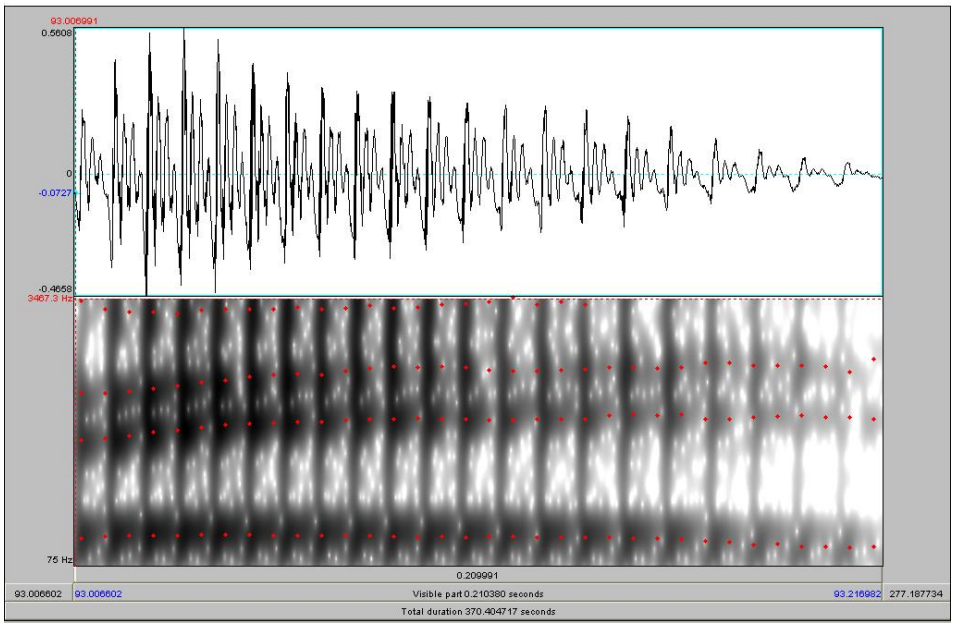
Osnovne akustičke karakteristike vokala mogu se videti u njihovim spektrima. Na slici 4.4 prikazani su spektrogrami svih pet vokala srpskog jezika. U akustičkim analizama ne samo govornog signala već i generalno zvuka, uobičajeno je signal predstavljati tzv. *spektrogramom*. Spektrogram omogućava trodimenzionalnu vizuelizaciju analiziranog signala. Na apscisi spektrograma nalazi se vreme, na ordinati frekvencija, a intenzitet signala predstavljen je (u presečnoj tački apscise i ordinate) intenzitetom zatamnjenja.



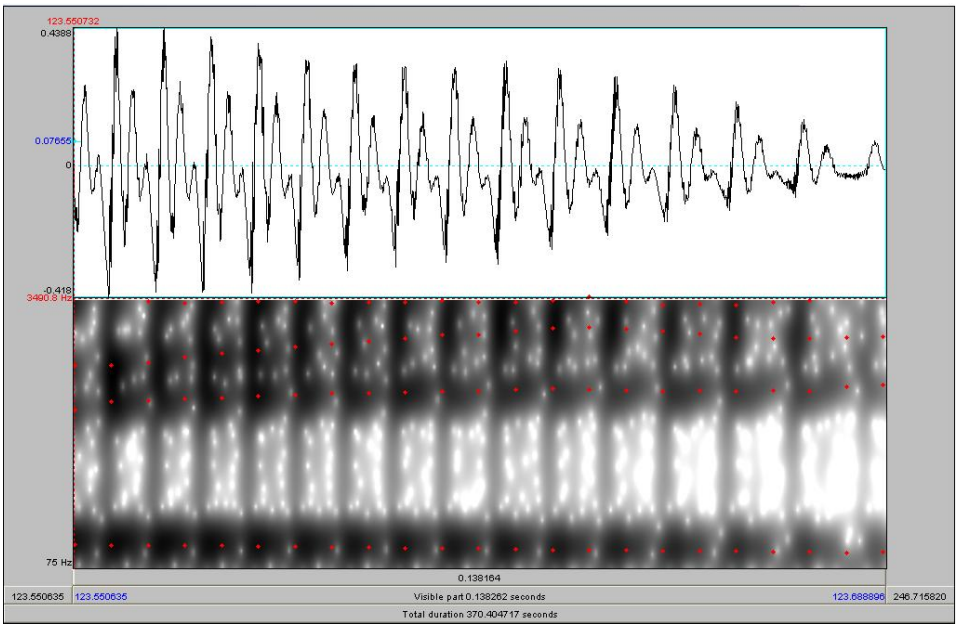
Slika 4.4.a – Talasni oblik i spektrogram vokala /a/

Slika 4.4 pokazuje da je spektar vokala harmonijske strukture. Najniži maksimum u spektru odgovara osnovnoj učestanosti govornog signala, koja se najčešće obeležava sa f_0. Ova

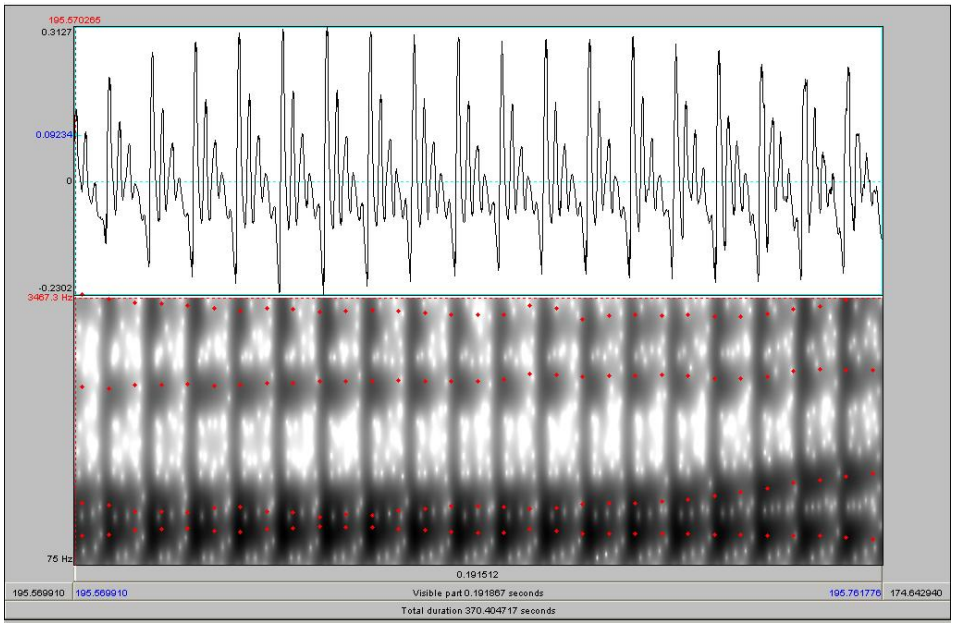
učestanost predstavlja učestanost ponavljanja laringealnih (glotalnih) impulsa. To je ono što mi subjektivno doživljavamo kao ”visinu” nečijeg glasa. Ova učestanost je kod muškog glasa najniža i nalazi se najčešće u granicama od 80 do 180 Hz, kod ženskog glasa ona je u intervalu od 180 do 230 Hz, dok je kod dečijeg glasa između 230 i 300 Hz (Vladisavljević, 1981).



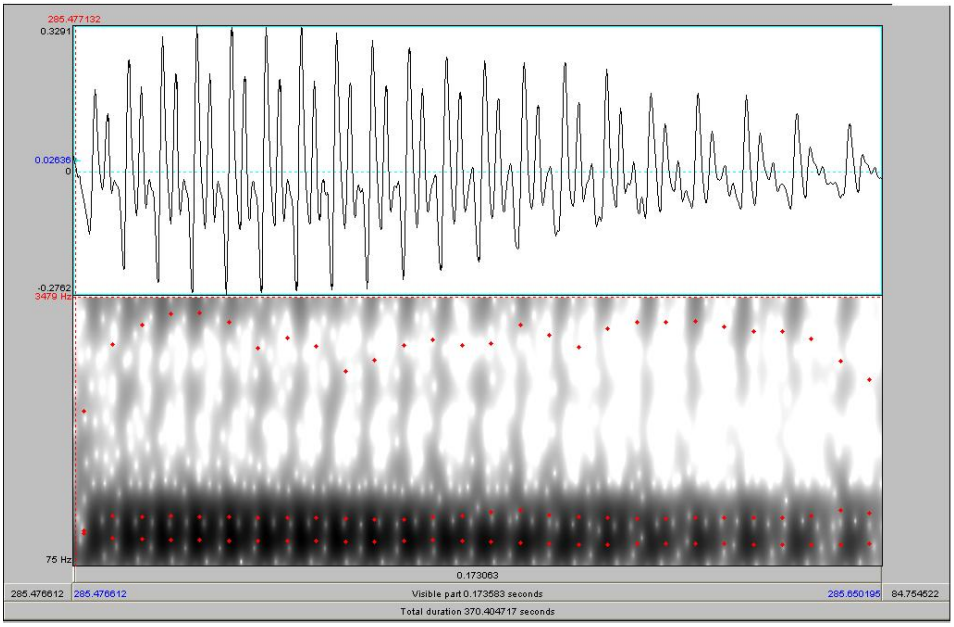
Slika 4.4.b – Talasni oblik i spektrogram vokala /e/



Slika 4.4.c – Talasni oblik i spektrogram vokala /i/



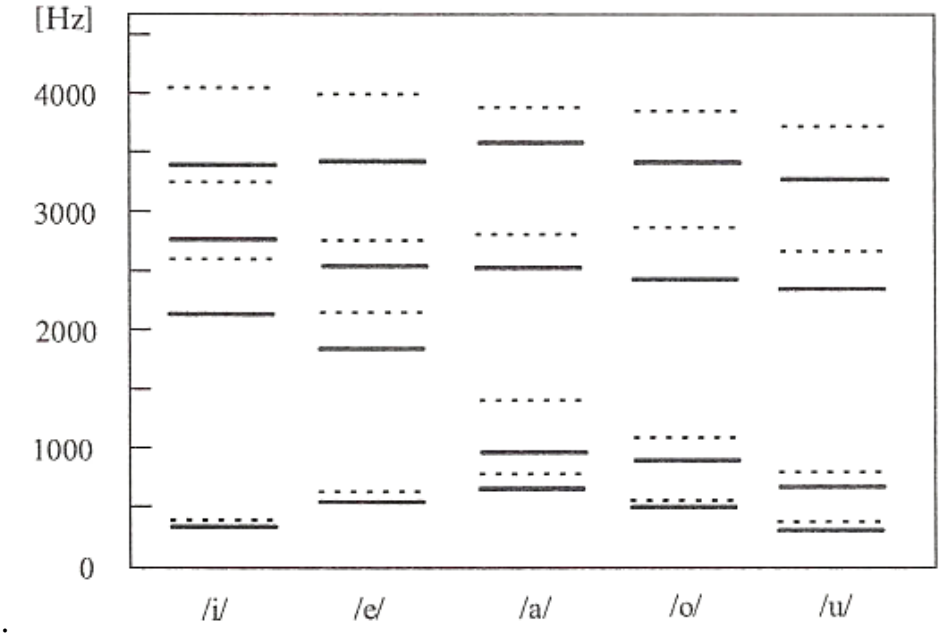
Slika 4.4.d – Talasni oblik i spektrogram vokala /o/



Slika 4.4.e – Talasni oblik i spektrogram vokala /u/

U proseku individualna visina glasa varira u okviru jedne oktave a zavisi od leksičkog akcenta, melodije rečenice i značenja koje se pridaje nekom iskazu. Modifikacijama osnovne frekvencije se pored značenja iskaza prenose i emocionalni efekti. Osnovna učestanost jeste individualna karakteristika govornika, koja može pomoći pri utvrđivanju emocionalnog ili zdravstvenog stanja govornika. Pored osnovne učestanosti u spektrima vokala ističu se i

određena spektralna područja. Ovi spektralni vrhovi nazivaju se formanti. U spektru vokala može biti do pet formanata označenih sa f_1, f_2, ... Prva tri formanta nose osnovna obeležja vokala, dok su za prepoznavanje vokala dovoljna samo prva dva formanta, f_1 i f_2. Treći formant f_3 daje jasnoću i poboljšava kvalitet glasa, on karakteriše labijalizaciju, odnosno učestvuje u razlikovanju nezaobljenih (nelabijalizovanih) i zaobljenih (labijalizovanih) glasova. Formanti su takođe individualne karakteristike na osnovu kojih se govornici mogu razlikovati. Međutim, na varijacije formanata utiču i koartikulacija susednih fonema u reči (ili rečenici) kao i akcenti. Na slici 4.5 (Jovičić, 1999) prikazan je raspored prva četiri formanta za muški i ženski glas. Kod ženskog glasa formanti su za oko 18 % viši u odnosu na muški glas (Kostić, 1971).



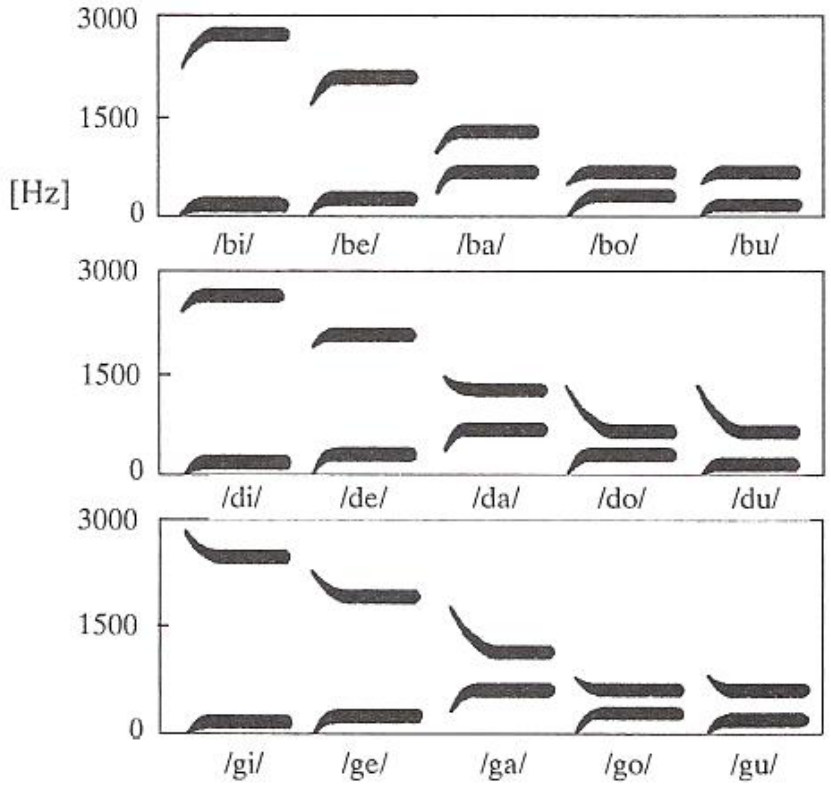
Slika 4.5 – Raspored formanata vokala za muški (___) i ženski (....) glas

Akustičke karakeristike koje su sa lingvističkog aspekta bitne za identifikaciju određenog fonema, kao i za međusobnu distinkciju fonema dele se na invarijantne i tranzicione karakteristike. Invarijantne karakteristike se definišu kao akustičke karakteristike koje pripadaju određenom fonemu a ne zavise od susednih fonema u datom kontekstu. Prema tome, one definišu identitet fonema bez obzira na to gde se on nalazi u govornom tekstu, ali to još uvek ne znači da je na osnovu invarijantnih karakteristika određeni fonem moguće apsolutno identifikovati. Tranzicione karakteristike su akustičke karakteristike koje imaju prelaznu formu najčešće između konsonanata i vokala ili obrnuto, i posledica su međusobnog uticaja fonema,

odnosno koartikulacije. Ove karakteristike su zavisne od konteksta i doprinose međusobnoj distinkciji fonema.

Invarijantne karakteristike vokala, na osnovu kojih se oni mogu razlikovati od konsonanata su čista zvučnost (bez prisustva frikcije), intenzitet, oblik spektra (postojanje formanata u spektru) i trajanje. Kao što je ranije napomenuto, za međusobnu distinkciju vokala neophodna su bar prva dva formanta.

Tranzicije formanata u spektru vokala, koji se u kontekstu nalaze ispred ili iza određenog konsonanta, predstavljaju najznačajnije tranzicione karakteristike. Na slici 4.6 (Jovičić, 1999) prikazane su tranzicije prvih i drugih formanata svih pet vokala u kontekstu zvučnih ploziva /b/, /d/ i /g/ na inicijalnoj poziciji.



Slika 4.6 – Tranzicije formanata vokala u kontekstu zvučnih ploziva /b/, /d/ i /g/ na početnoj poziciji

Tranzicije prvih formanata su uvek iste, stoga one očigledno ne utiču na razlikovanje ploziva. Tranzicije drugih formanata kod ploziva /b/, čije je mesto artikulacije u prednjem delu vokalnog trakta, uzlazne su i nezavisne od konteksta, odnosno susednih vokala. One se razlikuju po veličini, pre svega u frekvencijskom domenu. Kod ploziva /g/, čije je mesto artikulacije u

zadnjem delu vokalnog trakta, tranzicije su silazne. Mesto artikulacije ploziva /d/ je u srednjem delu vokalnog trakta i kod njega su tranzicije drugog formanta uzlazne ili silazne, u zavisnosti od vrste vokala. Veličina tranzicija se bitno razlikuje od sličnih tranzicija kod /b/ i /g/.

4.3 SUGLASNICI (KONSONANTI)

Suglasnički sistem srpskog jezika sastoji se od 25 lokalizovanih glasova koji se međusobno razlikuju po zvučnosti, mestu i načinu artikulacije. Osnovna podela koja se može izvršiti kod ove kategorije glasova jeste podela na sonante (glasnike) i konsonante (suglasnike). Pri artikulaciji sonanata prepreka je slaba tako da fonaciona struja nalazi prostora za slobodan prolaz kao pri artikulaciji samoglasnika. Sonanti su glasovi koje odlikuje kombinacija tona i šuma i po tome predstavljaju prelaznu klasu između vokala i konsonanata. Sonanti se, dakle, po svojim fiziološkim karakteristikama nalaze između samoglasnika i suglasnika, ali se ipak ubrajaju u suglasnike jer su lokalizovani, imaju mesto artikulacije. Zahvaljujući dvostrukim obeležjima artikulacije, slobodan prolaz fonacione struje, na jednoj, i postojanje pregrade na drugoj strani, sonanti najčešće ne podležu zakonu o jednačenju po zvučnosti, a neki od njih mogu se javiti i u ulozi nosioca sloga. Takođe, kao i vokali svi sonanti su zvučni glasovi. S obzirom na prirodu prepreke i na puteve kojima se fonaciona struja (vazdušna masa) usmerava prilikom artikulacije ove klase glasova, oni se mogu podeliti na usne (oralne) kod kojih se vazdušna masa usmerava kroz usni rezonator i nosne (nazalne) kod kojih se vazdušna masa usmerava kroz nos. U oralne sonante spadaju /l/, /lj/ i /r/, dok su /m/, /n/ i /nj/ nazalni sonanti. Većina autora glas /j/ u srpskom jeziku smatra sonantom (Petrović & Gudurić, 2010).

4.3.1 Artikulaciono-akustičke karakteristike konsonanata

Prema načinu artikulacije, odnosno prema tome kakva se prepreka stvara fonacionoj struji u usnoj duplji prilikom njihove artikulacije, suglasnici se mogu klasifikovati u nekoliko grupa:

1. plozivi (eksplozivni, praskavi, pregradni, zatvorni, okluzivi, prekidni)
   /p/, /t/, /k/, /b/, /d/, /g/
2. frikativi (strujni, trveni, spiranti, šumni)
   /f/, /h/, /s/, /š/, /v/, /z/, /ž/
3. afrikati (sliveni, semiokluzivi)

/c/, /ć/, /č/, /đ/, /dž/

4. nazali (nosni)

   /m/, /n/, /nj/

5. laterali (bočni)

   /l/, /lj/

6. vibrant

   /r/

7. poluvokal

   /j/

S obzirom na svoja akustička svojstva, nazali i laterali, kao i vibrant /r/ i poluvokal /j/ pripadaju ujedno i klasi sonanata.

Prilikom artikulacije ploziva govorni organi stvaraju potpunu pregradu koju fonaciona struja probije, pa nastaje mali prasak (eksplozija). Artikulacija ploziva sastoji se iz dva dela: period okluzije (trajanje pregrade pod pritiskom) i period eksplozije (probijanje pregrade), čemu prethode stvaranje pregrade i nagomilavanje vazdušne mase (implozija).

Pri artikulaciji frikativa vazdušna struja prolazi kroz sužen glasovni kanal tarući se o njegove zidove i stvarajući pri tome izraziti šum.

Pri artikulaciji afrikata prvo se stvara puna prepreka, koja se pretvara u tesnac pre nego što je fonaciona struja probije, tako da ne dolazi do prave eksplozije. Afrikati predstavljaju posebnu kategoriju glasova koji su sliveni od dva elementa, jednog praskavog i jednog strujnog. Oni se na perceptivnom planu doživljavaju kao jednoobrazni, ali zapravo predstavljaju kombinaciju dve različite artikulacije.

Prilikom artikulacije laterala jezik stvara pregradu po sredini krova usta, levi i desni tok fonacione struje prolazi nesmetano dok se oko pregrade stvara određeni nivo frikcije. Kod nazala pregrada se stvara u ustima ili na usnama a tok vazdušne struje usmeren je kroz nosne šupljine. Za vibrant /r/ karakteristični su prekidi fonacione struje oko kojih se stvara frikcija, dok između prekida fonaciona struja slobodno protiče stvarajući harmonijski oblik akustičke energije. Ukoliko se istovremeno pojave harmonijski i šumni oblik vazdušne struje tada se generiše poluvokal /j/.

U tabeli 4.3 (Jovičić, 1999) dato je artikulaciono polje konsonanata, koje obuhvata klasifikaciju konsonanata na osnovu tri osnovne distinktivne karakteristike: prema zvučnosti, prema mestu i načinu artikulacije. S obzirom na to da mesto artikulacije fonema u velikoj meri zavisi od konteksta, odnosno fonetskog okruženja, prilikom artikulacije jednog fonema često

dolazi do promene mesta njegove artikulacije, što ukazuje da se radi o području artikulacije koje obuhvata veći broj mesta artikulacije.

Artikulaciono polje	Plozivi		Afrikati		Frikativi		Nazali		Laterali		Vibrant		Poluvokal	
	zv	bzv	zv	bzv	zv	bzv	zv	bzv	zv	bzv	zv	bzv	zv	bzv
Bilabijalni	/b/	/p/			/v/		/m/							
Labio-dentalni					/v/	/f/								
Interdentalni	/d/	/t/		/c/	/z/ /ž/	/s/ /š/	/n/							
Dentalni	/d/	/t/		/c/	/z/ /ž/	/s/ /š/	/n/							
Postdentalni	/d/	/t/		/c/ /ć/	/z/ /ž/	/s/ /š/	/n/		/l/					
Postdentalno-alveolarni			/đ/	/c/ /ć/	/ž/	/š/	/n/		/l/					
Alveoralni			/đ/ /dž/	/ć/ /č/	/ž/	/š/			/l/ /lj/		/r/			
Alveo-palatalni			/dž/	/č/					/lj/					
Palato-alveolarni							/nj/						/j/	
Palatalni	/g/	/k/				/h/	/nj/						/j/	
Palato-velarni	/g/	/k/				/h/								
Velarno-palatalni	/g/	/k/				/h/								

Tabela 4.3 – Klasifikaciono polje konsonanata

4.3.1.1 Karakteristike ploziva

U grupu eksplozivnih suglasnika (ploziva) ubraja se šest fonema, tri bezvučna: /p/, /t/ i /k/ i tri zvučna: /b/, /d/ i /g/. Artikulacija ploziva sastoji se iz dva dela: perioda okluzije i perioda eksplozije. Period okluzije je znatno duži od perioda eksplozije. Tokom trajanja okluzije kod bezvučnih ploziva nastaje potpuna tišina, dok se kod zvučnih ploziva javljaju laringealne vibracije. Pri artikulaciji ploziva /p/ i /b/ prepreku fonacionoj struji u vokalnom traktu stvaraju usne. Dakle, prema mestu artikulacije oni su bilabijalni (dvousneni) plozivi, s tim da su kod artikulacije ploziva /b/ aktivne i glasne žice. Zbog toga se /p/ i /b/ nazivaju akustičkim parom. Pri izgovoru /t/ i /d/ pregradu prave jezik i zubi, te su oni stoga dentalni (zubni) plozivi. U izgovoru

/k/ i /g/ pregradu stvara zadnji deo jezika i zadnje (meko) nepce, te su oni velarni (zadnjonepčani) plozivi. Fonemi /t/ i /d/, kao i /k/ i /g/ predstavljaju akustički par.

Prilikom artikulacije ploziva /p/ i /b/ usne su zatvorene, jezik leži na dnu usne duplje stvarajući veliki oralni rezonator dok je velum podignut. Kada se vazdušna struja probije kroz pregradu, naglim spuštanjem donje vilice i razdvajanjem usana, čuje se prasak. Akustička energija se pojavljuje u vidu više frikcionih koncentrata u frekvencijskom opsegu od 300 do 5500 Hz. Kod artikulacije zvučnih fonema u toku okluzije glasne žice su aktivne smanjujući tenziju artikulacionih organa. Generalno, svi zvučni plozivi su manjeg intenziteta od bezvučnih.

Pri izgovoru ploziva /t/ i /d/ pregrada se stvara tako što se vrh jezika jako pribija uz gornje sekutiće, gornja površina jezika na alveole, a bokovi na gornje zube. Tako se stvara mali rezonator u prednjem oralnom prostoru usled čega se eksplozija karakteriše šumnom energijom koja se u vidu više koncentrata gotovo ravnomerno rasprostire do oko 8000 Hz. Pregradu vazdušnoj struji kod ploziva /k/ i /g/ stvara zadnji deo jezika koji se podiže do mekog nepca, dok vrh jezika leži na dnu usne duplje iza donjih sekutića. Zbog ovakvog kretanja jezika unazad stvara se veći rezonator u prednjem oralnom prostoru koji nakon eksplozije omogućava stvaranje šumne energije u vidu dva šira koncentrata, i to intenzivnijeg oko 2000 Hz i slabijeg oko 5000 Hz.

Dužina trajanja okluzije i eksplozije zavisi od pokretljivosti najaktivnijih artikulacionih organa u stvaranju pregrade. Usne su najpokretljivije, zatim vrh jezika a najsporiji je zadnji deo jezika. Stoga, bilabijalni plozivi traju najkraće, a velarni (zadnjonepčani) najduže.

4.3.1.2 Karakteristike frikativa

Strujni suglanici (frikativi) nastaju trenjem vazdušne stuje o zidove tesnaca koji stvaraju pokretni govorni organi (jezik i usne) sa nepokretnim (gornji sekutići, nepce). Fonetsku grupu frikativa čine četiri bezvučna fonema: /f/, /s/, /š/ i /h/ i tri zvučna fonema: /v/, /z/ i /ž/. Frikativi /f/ i /v/ obrazuju se uz tesnac koji prave donja usna i gornji sekutići, te su stoga oni labijalno-dentalni (usneno-zubni) frikativi. Pri izgovoru glasova /z/ i /s/ tesnac grade jezik i gornji zubi, pa su oni dentalni (zubni) frikativi. Frikativi /ž/ i /š/ su alveolarni (nadzubni) frikativi jer nastaju podizanjem jezika u pravcu alveolarnog ruba i povlačenjem unazad čime se povećava prednji alveolarni prostor, dok je /h/ velarni (zadnjonepčani) frikativ koji se obrazuje s tesnacem koji gradi zadnji deo jezika sa zadnjim (mekim) nepcem.

Labijalno-dentalni frikativi /f/ i /v/ obrazuju se tako što se fonaciona struja probije kroz tesnac koji prave gornji sekutići i unutrašnja, vlažna, strana donje usne. Kod fonema /f/ akustička

energija šumne prirode rasprostire se u vrlo širokom opsegu od 500 do blizu 10000 Hz u blago naglašenim koncentratima. Kod fonema /v/ javlja se laringealna zvučnost koja u spektru formira više formantnih ravnomerno raspoređenih koncentrata. Zbog svoje poluformantne strukture fonem /v/ se ponekad svrstava u sonante (Jovičić, 1999).

Konstrikciju dentalnih fonema /s/ i /z/ stvaraju gornji i donji sekutići i vrh jezika koji se podiže ka gornjim sekutićima ali ne dodiruje postdentalni prostor. Spektar frikcije nalazi se u gornjem frekvencijskom području od 3000 do blizu 10000 Hz sa slabim koncentratom od 6000 do 9000 Hz. Frikativ /z/ je zvučni parnjak bezvučnom /s/.

Kod artikulacije alveolarnih frikativa /š/ i /ž/ prednji deo jezika podiže se u pravcu alveolarnog ruba i povlači nešto unazad, usled čega se povećava prednji alveolarni prostor. Frikcija se stvara neposredno iza vrha jezika a zbog uvećanog alveolarnog prostora šumni spektar je kod ovih fonema, u poređenju sa dentalnim fonemima, niži i intenzivniji. Kod fonema /š/ i /ž/ spektar se sastoji od dva široka i izrazita koncentrata frikcione energije. Prvi je ispod 2000 Hz do blizu 4000 Hz, a drugi od 4000 do iznad 7000 Hz.

Kod artikulacije fonema /h/ vrh jezika se povlači od donjih sekutića i leži na dnu usne duplje, čime se stvara veliki prednji oralni prostor koji kod frikcije doprinosi isticanju šumnog spektra od 200 Hz do blizu 4500 Hz koji je prilično ravnomerno raspoređen bez izrazitih koncentrata, ali sa blago naglašenom poluformantnom strukturom.

4.3.1.3 Karakteristike afrikata

Sliveni suglasnici (afrikati) nastaju kombinacijom artikulacionih pokreta karakterističnih za artikulaciju ploziva i frikativa. Afrikati nastaju otvaranjem pregrade, ali se ona ne otvara naglo. Postepenim otvaranjem pregrade eksplozija slabi i dolazi do pojačane šumnosti, afrikcije. Prasak afrikcije (plozija) podseća na trenutak eksplozije kod ploziva. Nakon plozije sledi kratka pauza koja odvaja ploziju od afrikcije. Fonetsku grupu afrikata čine tri bezvučna fonema /c/, /ć/ i /č/ i dva zvučna fonema /đ/ i /dž/.

Kod artikulacije afrikata /c/ vrh jezika stvara pregradu odmah iza gornjih sekutića, dok se bočne strane jezika naslanjaju na desni i kutnjake gornje vilice. Fonaciona struja koja dolazi iz farinksa usmerava se kroz žleb koji se stvara sredinom jezika sve do vrha jezika gde počinje pregrada da se otvara. Tu se stvara mali rezonator u kome se fonaciona struja pojačava i koja se konačno probija kroz međuzubni i međuusni prostor. Ovaj rezonator karakteriše šumnu energiju fonema /c/ u vrlo visokom frekvencijskom području do iznad 9000 Hz. Afrikat /c/ je po mestu artikulacije dentalni (zubni) suglasnik.

Kod afrikata /ć/ i /đ/ vrh jezika stvara prepreku u postdentalnom prostoru, dok se prednji i srednji delovi jezika priljubljuju uz nepce. Vrh jezika je na unutrašnjoj strani donjih sekutića dok je vilični ugao nešto povećan. Usled toga je prednji oralni rezonator uvećan u odnosu na fonem /c/, pa se najintenzivniji deo šumnog spektra afrikcije stvara u opsegu od 3000 do 5000 Hz. Afrikcija zauzima široko spektralno područje od 2000 do oko 9000 Hz. Prema mestu artikulacije ovi fonemi su postdentalni (zazubni). Afrikat /đ/ je zvučni parnjak bezvučnom /ć/.

Kod artikulacije afrikata /č/ i /dž/ pregrada se najčešće stvara vrhom jezika i alveolarnog ruba. Strane jezika se naslanjaju na krov usta i kutnjake. Sa ovakvim položajem jezika stvaraju se dva relativno velika rezonatora, iza pregrade prema farinksu i ispred pregrade između usana i donje površine jezika. Prednji rezonator je uvećan usnim rezonatorom jer su pri artikulaciji fonema /č/ i /dž/ usne zaobljene i malo isturene napred. Između plozije i afrikcije nema izrazite pauze, najizrazitiji koncentrat afrikcije zauzima opseg od 3000 do oko 6000 Hz. Prema mestu artikulacije reč je o alveolarnim fonemima, afrikat /dž/ je zvučni parnjak bezvučnom /č/.

4.3.1.4 Karakteristike nazala

Nosni suglasnici (nazali) su oni suglasnici pri čijoj artikulaciji fonaciona struja prolazi kroz dva različita rezonatora: usnu duplju i nosnu šupljinu. Fonetsku grupu nazala čine fonemi /m/, /n/ i /nj/. Stvaranje nazalnog suglasnika podrazumeva stvaranje prepreke u ustima ili na usnama, slobodno prolaženje vazduha kroz nosne šupljine i pored stvorene prepreke a u trenutku kada se prepreka ukloni vazduh istovremeno prolazi kroz oba rezonatora. Slobodno kretanje vazduha kroz nosnu šupljinu usloviće pojavu formanata na čije će karakteristike u najvećoj meri uticati vibracije nastale u nazalnom delu glasovnog kanala i u farinksu. Na spektrogramima je uočljivo prisustvo formanata u gornjim frekvencijama za fonem /m/, slabije prisustvo kod fonema /n/ dok kod fonema /nj/ nema jasno vidljivih formanata u gornjim frekvencijama.

Fonem /m/ je bilabijalni (dvousneni) nosni (nazalni) suglasnik. Prilikom artikulacije ovog fonema usne su celom dužinom priljubljene jedna uz drugu, stvarajući tako prepreku. Fonaciona struja počinje da teče kroz nos, kuda joj je omogućen prolaz time što je zadnje (meko) nepce spušteno, a kad se usne razmaknu, ostatak vazdušne struje prođe i kroz usta.

Fonem /n/ je dentalni (zubni) nazalni suglasnik. Izgovara se tako što je vrh jezika pritisnut uz alveole gornjih sekutića a meko nepce se spušta, otvarajući prolaz vazdušnoj struji kroz nos. Po otklanjanju jezika sa alveola deo vazdušne struje prolazi i kroz usne.

Kod artikulacije fonema /nj/ postoji potpuna pregrada prolasku fonacione struje u usnoj duplji, ali joj je put kroz nos potpuno slobodan. Vrh jezika nalazi se uz donje sekutiće, a gornja

površina uz prednje (tvrdo) nepce. Prema mestu artikulacije ovaj fonem spada u palatalne suglasnike. Kada se pregrada otvori, deo fonacione struje prolazi kroz usta.

Najjače izražen nazalni formant nalazi se na frekvenciji od oko 250 Hz, dok se drugi formant obično formira na visini od oko 2500 Hz. Pojas između 500 i 2000 Hz je, u načelu, prigušen, ali i tu postoje razlike u zavisnosti od toga koji je nazalni suglasnik u pitanju. Prigušenje je najizraženije kod fonema /nj/, zatim kod /n/, dok je kod /m/ ono najmanje izraženo.

4.3.1.5 Karakteristike laterala

U bočne suglasnike (laterale) ubrajaju se fonemi /l/ i /lj/. Prilikom artikulacije fonema /l/ vrh jezika dodiruje gornji deo vilice najčešće u predelu alveola, mada može dodirivati i zube. Ukoliko se leđa jezika izdignu dodirujući palatinalni luk, a pri tom vazduh slobodno prolazi sa obe strane jezika (lateralno) artikulisaće se glas /lj/.

Artikulacija laterala sastoji se od dve faze. Prvu fazu čini period trajanja laterala pre otvaranja pregrade. U ovoj fazi laterale karakterišu četiri poluformantna koncentrata akustičke energije. Prvi do 1000 Hz, treći od oko 3000 do 3500 Hz i četvrti od 4500 do 5000 Hz su praktično isti poluformanti za oba laterala. Vidljiva razlika postoji u položaju drugog poluformanta. Kod /l/ drugi poluformant se nalazi u opsegu od 1000 do 1500 Hz, dok je kod /lj/ on u opsegu od 2500 do 3000 Hz. Ovu fazu karakteriše kombinacija zvučne i frikcione akustičke energije. Nakon otvaranja pregrade nastaje druga faza u artikulaciji laterala. U zavisnosti od fonema koji sledi nakon laterala ova faza se jasno karakteriše frikcionim oblikom akustičke energije gde se gubi poluformantna struktura.

4.3.1.6 Karakteristike vibranta

U srpskom jeziku fonem /r/ je jedini vibracioni fonem. Nastaje vibracijom prednjeg dela jezika tako što vrh jezika u potunosti naleže na alveolarni rub stvarajući pregradu. Ova okluzija traje manje od 20 ms. Pod pritiskom vazdušne struje pregrada se otvara i u kratkotrajnoj eksploziji čuju se vibracije glasnica. Ovaj zvučni period je poluformantnog karaktera zahvaljujući velikim rezonantnim prostorima između pregrade i larinksa. Inervacija mišića jezika vrlo brzo vraća vrh jezika u prethodni položaj stvarajući ponovo pregradu, pri čemu dolazi do vibriranja. Pri artikulaciji fonema /r/ broj vibracija je od 1 do 4-5 u zavisnosti od konteksta, naglašavanja i načina artikulacije. Prema mestu artikulacije fonem /r/ je u srpskom jeziku alveolarni fonem.

U fazi zvučnosti, nakon eksplozije, jasno se uočavaju frikcioni i formantni periodi. Početni i završni deo ove faze su frikcionog karaktera dok je središnji deo dominantno formantne strukture. Poluformantni koncentrati akustičke energije izraženi su u opsezima do 1000 Hz, od 1500 do 2000 Hz i oko 3000 Hz.

4.3.1.7 Karakteristike poluvokala

Artikulacija fonema /j/ vrlo je slična artikulaciji vokala /i/ zbog čega se u klasifikaciji fonema sonant /j/ naziva poluvokalom. Pri artikulaciji fonema /j/ jezik se iz položaja za vokal /i/ podiže naviše i pomera napred tako da svojim stranama dodiruje kutnjake stvarajući po sredini levak između središnjeg dela jezika i tvrdog nepca. Stepen frikcije koji će biti izražen u akustičkoj strukturi fonema /j/ zavisi od veličine ovog levka. Osobine fonema /j/ umnogome zavise od koartikulacije. Na početku reči fonem /j/ je uvek naglašen sa vidno prisutnom frikcionom komponentom. U finalnom položaju, kada je nenaglašen, frikciona komponenta iščezava, slično kao kada je u položaju između vokala kada dobija osobine poluvokala.

U akustičkoj strukturi fonema /j/ izražena su četiri poluformantna koncentrata akustičke energije, prvi od oko 100 do 400 Hz, drugi oko 2500 Hz, treći iznad 3000 Hz i četvrti iznad 4000 Hz. Prva tri koncentrata odgovaraju položajima formanata vokala /i/, što potvrđuje konstataciju o velikoj sličnosti sonanta /j/ i vokala /i/.

4.4 ZNAČAJ AKUSTIČKIH KARAKTERISTIKA U PERCEPCIJI VOKALA

Osnovna učestanost, formanti, tranzicije osnovne učestanosti i formanata, kao i trajanje vokala jesu osnovne akustičke karaktersitike na osnovu kojih se vokali međusobno razlikuju. Međutim, značaj određene akustičke karakteristike u percepciji vokala nije isti za sve vokale svih jezika (Jovičić, 1999). Tako na primer, u kineskom i tajlandskom jeziku, koji spadaju u grupu tonskih jezika, vokali istih formanata ali različitih tranzicija osnovne učestanosti percipiraju se kao različiti fonemi. Takođe, u srpskom jeziku postoje reči koje promenom akcenta menjaju svoje značenje.

Analiza uticaja osnovne učestanosti i formanata na percepciju vokala u eksperimentima čiji se rezultati mogu pronaći u literaturi vršena je na osnovu ocene kvaliteta vokala (Jovičić, 1999). Pre svega, psihoakustički testovi su pokazali da osnovna učestanost igra sekundarnu ulogu u percepciji vokala, kao i da spektralni sastav vokala ne utiče na percepciju osnovne

učestanosti (Schouten et al., 1962). Potom, pokazalo se da sinhrono pomeranje svih formanata za isti procenat duž frekvencijske skale, ka višim ili nižim frekvencijama, vrlo malo utiče na identifikaciju vokala (Peterson & Barney, 1952). Evidentan primer jeste percepcija vokala kod muškog, ženskog i dečijeg glasa. Takođe, ukoliko se osnovna učestanost udvostruči, tada je za očuvanje kvaliteta vokala neophodno formante pomeriti za 15 % ka višim učestanostima (Slawson, 1968).

Zbog koartikulacije sa susednim fonemima u kontinualnom govoru formanti vokala najčešće ne dostižu vrednosti koje imaju kod izolovanog izgovora. Međutim, ova činjenica ne utiče na identifikaciju vokala u nekom kontekstu. U literaturi je ova činjenica poznata pod nazivom ”redukcija vokala” (Lindblom, 1963).

Pored prvog i drugog formanta koji predstavljaju primarne akustičke karakteristike u prepoznavanju vokala, pokazalo se da je trajanje vokala treća karakteristika od važnosti za percepciju vokala (Cohen et al., 1967). Činjenica je da su vokali najpodložniji promeni trajanja usled promene tempa govora, ali postoje i mnogi drugi faktori koji utiču na trajanje vokala, o čemu je već bilo reči u prethodnim poglavljima.

Prvi i drugi formant predstavljaju primarne akustičke karakteristike u identifikaciji vokala. Međutim, vrednosti formanata znatno variraju između različitih govornika, tako da se polja varijacije vokala preklapaju. S druge strane, slušalac ipak dobro prepoznaje vokale, što ukazuje na činjenicu da se identifikacija vokala ne može vršiti samo na osnovu apsolutnih vrednosti formanata, već da se na određeni način mora izvršiti normalizacija vrednosti formanata, odnosno normalizacija vokalnog trakta. Stoga, potrebno je pronaći određenu međuzavisnost između formanata ili drugih akustičkih karakteristika i ona treba da je invarijantna u odnosu na govornike. Uočena korelacija između osnovne učestanosti i formanata ukazuje da bi osnovna učestanost mogla biti normalizujući faktor (Jovičić, 1999). Problem normalizacije vokalnog trakta je od interesa u sistemima za prepoznavanje govora, kao i u sistemima konverzije jednog glasa u drugi (Choi & King, 1995).

4.5 ZNAČAJ AKUSTIČKIH KARAKTERISTIKA U PERCEPCIJI KONSONANATA

Skup akustičkih karakteristika koje opisuju određeni konsonant mnogo je širi i kompleksniji nego kod vokala, jer artikulacija konsonanata nastaje znatno bržim pokretima artikulatora i većim konstrikcijama vokalnog trakta. Multidimenzionalnom analizom utvrđeno je da se najveći broj konsonanata može identifikovati pomoću tri distinktivne karakteristike: prema

načinu artikulacije, zvučnosti i prema mestu artikulacije. Takođe je ustanovljeno da način artikulacije vrlo precizno pravi distinkciju između grupa fonema. Detaljnija analiza akustičkih karakteristika pokazala je da se u identifikaciji načina artikulacije najveće razlike pojavljuju kod karakteristika plozivnosti, kontinualnog frikativnog šuma i nazalnih rezonancija (Jovičić, 1999). Prema načinu artikulacije konsonanti u srpskom jeziku dele se u nekoliko grupa: plozivi, frikativi, afrikati, laterali, nazali, vibrant i poluvokal.

Karakteristika zvučnosti podrazumeva postojanje laringealnih vibracija prilikom artikulacije zvučnih konsonanata. Plozivi, frikativi i afrikati imaju zvučne i bezvučne foneme. Istraživanja su pokazala da su osnovne akustičke karakteristike koje omogućavaju perceptivno identifikovanje zvučnosti konsonanata: trajanje vremenskog intervala od eksplozije ploziva do početka vokalnih vibracija – VOT (engl. *Voice Onset Time*), brzina promene tranzicija formanata vokala koji slede iza konsonanata, aspiracija kod bezvučnih ploziva, trajanje vokala koji prethode plozivima i frikativima i spektralni sastav eksplozivnog praska kod ploziva. Takođe, utvrđeno je da zvučnost ploziva definišu kraći interval VOT i veće tranzicije formanata (Jovičić, 1999). U mnogim istraživanjima primećeno je da na identifikaciju zvučnosti konsonanata ne utiču samo karakteristike konsonanata već i karakteristike vokala koji im prethode (Jovičić, 1999). Merenjem trajanja vokala ispred konsonanata u srpskom jeziku utvrđeno je da vokali traju duže ispred zvučnih nego ispred bezvučnih konsonanata (Sovilj-Nikić, 2007), što ukazuje na činjenicu da se trajanje vokala ispred konsonanta može smatrati dovoljnom akustičkom karakteristikom za percepciju zvučnosti.

Mesto artikulacije konsonanata jeste artikulaciona karakteristika koja podrazumeva mesto u okviru vokalnog trakta gde se fonaciona struja najintenzivnije modifikuje, što se ostvaruje potpunim blokiranjem ili propuštanjem kroz tesnace. Svako mesto artikulacije odražava se u akustičkim karakteristikama generisanog konsonanta, što se može identifikovati u invarijantnim ili tranzicionim karakteristikama. Tranzicije drugog formanta u spektru vokala koji sledi nakon ploziva imaju primarnu ulogu u diskriminaciji ploziva, a samim tim i u identifikaciji mesta artikulacije. Međutim, na osnovu sprovedenih istraživanja ustanovljeno je da se na bazi tranzicija formanata može izvršiti distinkcija ploziva ali da za apsolutnu identifikaciju mesta artikulacije nisu dovoljne samo tranzicije formanata (Sharf & Hemeyer, 1972). Takođe je na osnovu nezavisnog ispitivanja uticaja nazalnih rezonancija i tranzicija formanata pokazano da nijedna od ove dve akustičke karakteristike nije niti invarijantna niti dovoljna za percepciju mesta artikulacije (Malecot, 1956). Pored tranzicija formanata, eksplozivni prasak i aspiracija ploziva su takođe akustičke karakteristike koje značajno doprinose identifikaciji i diskriminacija ploziva. Eksperimenti su pokazali da eksplozivni prasak i aspiracija nisu invarijantne akustičke

karakteristike i da su za apsolutnu identifikaciju ploziva neophodne i tranzicije pripadajućih vokala koji im slede.

Osnovni načini artikulacije su vokalnost, plozivnost, frikativnost, nazalnost i vibrantnost, pri čemu je afrikativnost kombinacija plozivnosti i frikativnosti, dok je vibrantnost fonema /r/ jedinstvena invarijantna akustička karakteristika. Artikulacija ploziva se ne može izvršiti ako se vokalni trakt ne zatvori za određeni period vremena, odnosno ako se ne stvori interval tišine. Stoga, interval tišine na određenom mestu govorne poruke ima fonetsko značenje (Liberman & Studdert-Kennedy, 1979). Odnosno, interval tišine slušaocu naznačava artikulacioni pokret govornika o zatvaranju vokalnog trakta i stvaranja okluzije. Artikulacija frikativnosti podrazumeva stvaranje tesnaca duž vokalnog trakta koji, prolaskom fonacione struje, stavaraju šum određenog spektralnog sastava. Percepcija spektralnih kvaliteta ovog šuma je osnova u identifikaciji frikativa. Takođe je u eksperimentima pokazano da je intenzitet frikativa vrlo bitna akustička karakteristika koja doprinosi distinkciji frikativa (Jovičić, 1999).

4.6 AKCENAT SRPSKOG STANDARDNOG JEZIKA

Akcenat ili naglasak predstavlja naročito isticanje jačine i visine određenog sloga u reči ili jednosložne reči u rečenici.

Slogovi su najmanje glasovne jedinice izgovorene jedinstvenom artikulacionom aktivnošću. Nosioci sloga, ili slogovna jezgra, najčešće su vokali a ređe sonanti (/r/, /l/, /n/). Međutim, treba napomenuti da ukoliko sonant preuzima ulogu nosioca sloga, tada se pored njega uvek javlja poluglas /ə/ ili šva (nem. *schwa*) koji zapravo nosi sva prozodijska obeležja i predstavlja nosioca sloga. Slogovi mogu biti različitih struktura, odnosno mogu imati različite kombinacije vokala V i konsonanata C. U srpskom jeziku preko 90 % čine slogovi tipa CV, CCV, CVC, V (Jovičić, 1999).

U srpskom jeziku je pored položaja naglašenog sloga bitna i vrsta akcenta, koja određuje relativni položaj visine glasa u naglašenom slogu u odnosu na ostale slogove, kao i trajanje sloga, i kao takva neretko utiče na značenje same reči. U našem standardnom jeziku postoje četiri vrste akcenata:

- kratkosilazni – kao u rečima *kȕća, brȁt, prȉjatelj, čȅp, tȍp*
- dugosilazni – kao u rečima *sûnce, sât, sîn, pêć,rôg*
- kratkouzlazni – kao u rečima *vòda, širìna, kùtija, stàklo, žèna*
- dugouzlazni – kao u rečima *rúka, ráditi, zíma, móda, pégla*.

Slog sa kratkosilaznim akcentom izgovara se kratko, jačina i visina tona naglo i jednovremeno padaju. Kod reči sa dugosilaznim akcentom naglašen slog se izgovara dugo. Ton je u početku visok, a zatim se istovremeno sa jačinom izgovora upadljivo spušta. Kod silaznih akcenata, sledeći slog počinje tonom koji je niži od tona na kraju akcentovanog sloga.

Slog sa kratkouzlaznim akcentom izgovara se kratko, visina tona raste, a jačina izgovora opada pre završetka izgovora ovako naglašenog sloga. Slog sa dugouzlaznim akcentom izgovara se dugo, a ton je pretežno ravan. Kod uzlaznih akcenata sledeći slog počinje tonom koji je isti ili viši od tona na kraju akcentovanog sloga, a zatim sledi pad tona.

Kao što akcentovane jednosložne reči i akcentovani slogovi u višesložnim rečima mogu biti po trajanju (kvantitetu) kratki i dugi, tako i neakcentovani slogovi u srpskom jeziku mogu biti kratki i dugi. Kratak može biti bilo koji slog ispred ili iza akcentovanog sloga (*pȅsma, sênka, telefonírati*). Međutim, dugi neakcentovan slog može se javiti samo iza naglašenog sloga – kao u rečima *bolèsnīk, jùnāk, dèšnjāk*. Ovaj prozodijski element naziva se *posleakcenatska* (*neakcentovana*) *dužina* i predstavlja produženi izgovor sloga koji sledi iza naglašenog sloga. U jednoj reči može biti jedna, dve pa i tri neakcentovane dužine (vòjnīk, vrâbācā, polìcājācā). Posleakcenatske dužine doprinose čujnosti zadnjih slogova u rečima, koji su u tom pogledu ugroženi slabljenjem fonacione struje (Telebak, 2011). Pored toga, neakcentovana dužina kao "peti akcenat" u srpskom jeziku daje našem jeziku poseban kvalitet i ima važnu ulogu u metrici. Takođe, ona je ponekad nosilac fonološke opozicije, odnosno doprinosi označavanju razlike u značenju reči kao u slučaju *kàmēn* (imenica) – *kàmen* (pridev), *ùlica* (jednina) – *ùlīcā* (množina) itd. Uprkos navedenim funkcijama, posleakcenatske dužine se u savremenom govoru sve više skraćuju i redukuju do gubljenja, što je naročito izraženo u gradskim sredinama i češće karakteristike govora kod mlađih osoba. Ukoliko u reči postoji više neakcentovanih dužina, prva od njih je najstabilnija. Takođe, posleakcenatska dužina je stabilnija neposredno iza akcenta nego dalje od njega (*vòjnīk, sȁobraćāj, nástavnīk*), kao i iza uzlaznih nego silaznih akcenata. Stoga, iako u teoriji ona ne mora biti neposredno iza u savremenom govoru je gotovo isključivo neposredno iza, i to uvek kratkouzlaznog akcenta (*lùdāk, čùdāk*), što danas predstavlja jedinu poziciju posleakcenatske dužine koja se zadržala u šumadijsko-vojvođanskom dijalektu.

Posleakcenatska dužina se obavezno javlja kod određenih oblika reči (Telebak, 2011):

- genitiva jednine nekih i genitiva množine svih imenica: *iz knj̏gē, do kȕćē itd., bez kûćā, preko pôljā, malo nȍvācā* itd.;
- instumentala jednine imenica ženskog roda: *òlōvkōm, rúkōm, nògōm* itd.;
- komparativa i superlativa prideva: *nâjvećī, nâjlepšā* itd.;

- određenog pridevskog vida: *nȍvī, nȍvā, nȍvū* itd., *stârī, stârā* itd.;
- nekih glagolskih oblika: prezenta (*čìtām, pîšēš,* itd.); imperfekta (*písāh, čȉtāše*, itd.); glagolskih priloga – sadašnjeg i prošlog (*rádēći, čìtajūći* itd.; *urádīvši, pročìtāvši* itd.), glagolskih prideva – radnog i trpnog (*ùzēla, pȍčēli* itd.; *ùzēt, prȍčitān* itd.);
- u nekim nastavcima za izvođenje reči: *-āč, ār* i sl. (*pèvāč, lȉmār* itd.).

Višesložne reči u srpskom jeziku mogu imati bilo koji od četiri akcenta, s tim što se silazni akcenat sme javiti samo na prvom slogu. Uzlazni akcenat može biti na bilo kom slogu višesložne reči, osim na poslednjem. Jednosložne naglašene reči mogu imati samo silazne akcente. Silazni akcenti su izvorni (predbalkanski) akcenti, dok su uzlazni (balkanski) nastali pomeranjem silaznih akcenata s kraja reči za jedan slog ka početku, pri čemu su menjali intonaciju odnosno postajali uzlazni (*junâk>jùnāk, livàda>lìvāda*, itd.). Na taj način je negde u XV veku od monotonske dvoakcenatske akcentuacije nastao standardni, politonski četvoroakcenatski sistem (Telebak, 2011).

Ukoliko se silazni akcenat nalazio na prvom slogu, on nije imao mogućnost da se pomeri, čime se i objašnjava vezanost silaznih akcenata za prvi slog. Kod reči koje nastaju dodavanjem prefiksa prostim glagolima sa silaznim akcentom dolazi do pomeranja akcenta za jedan slog, kako silazni akcenat ne bi ostao na unutrašnjem slogu, uz promenu intonacije, odnosno transformacije akcenta u kratkouzlazni (*stȁti – pòstati, dȁti – pròdati, prâvdati – òprāvdati, jêmčiti – zàjēmčiti, skâčēm – ùskāčēm, râdīš – ìzrādīš*). Na slogu sa koga je pomeren silazni akcenat ostaje njegov trag, odnosno posleakcenatska dužina. Međutim, treba napomenuti da se u savremenom govoru, naročito mladih, sve češće javljaju odstupanja od navedenih pravila akcentovanja, tako da se mogu čuti reči, uglavnom stranog porekla, izgovorene sa naglaskom na poslednjem slogu (*asistȅnt, dirigȅnt, lavabô, bifê, renomê*), kao i reči kod kojih se silazni akcenat nalazi na nekom od unutrašnjih slogova (*radijâtor, televîzija, recitâtor, kompozîtir, audîcija*). Dakle, u pitanju su reči stranog porekla kod kojih nije izvršena akcenatska adaptacija, odnosno silazni akcenat se nalazi izvan prvog sloga, što je veoma čest slučaj u zapadnim jezicima odakle se danas i preuzima najveći broj reči. Silazni akcenti izvan prvog sloga sve više se ustaljuju i u genitivu množine nekih domaćih imenica (*podâtākā, zadâtākā, domaćînstāvā, policâjācā, Palestînācā*). Takođe, u nekim složenicama domaćeg porekla javljaju se silazni akcenti na unutrašnjim slogovima (*primoprȅdaja, kupoprȍdaja, poljoprȉvreda*). U ovom slučaju zadržani su akcenti druge imenice u složenicama (*prȍdaja, prȅdaja, prȉvreda*) gde su silazni akcenti na prvom slogu u potpunosti regularni.

Svaka naglašena reč u srpskom jeziku ima jedan naglašen slog, osim superlativa određenih prideva i priloga (*nâjpotrèbnijī, nâjpopulàrnijī, nâjjednostàvnijī*), kao i određenih složenica (*bȉosféra, mȉkroklíma, àgroindùstrija, àviomehàničār, kȍntrarevolúcija, sȅveroìstočni*), koje mogu imati i dva naglašena sloga. Pored naglašenih (akcentogenih) reči, u srpskom jeziku postoje i reči bez akcenta – neakcentogene reči. One se nazivaju *klitike* i izgovaraju se zajedno sa naglašenim rečima, odnosno sa njima čine akcenatsku celinu. Nenaglašene reči koje čine akcenatsku celinu sa rečju iza sebe nazivaju se *proklitike*. U ovu grupu reči spada većina predloga, veznika, kao i odrična rečca *ne*. *Enklitike* su nenaglašene reči koje stoje iza naglašenih reči i sa njima se zajedno izgovaraju. To su obično nenaglašeni oblici ličnih zamenica, kraći oblici pomoćnih glagola i upitna rečca *li*.

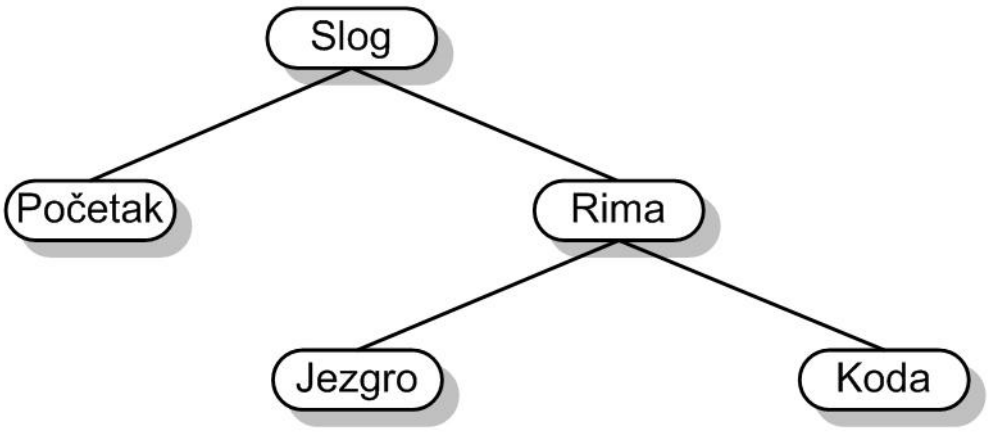
Kao što je već napomenuto, proklitike čine izgovornu celinu sa narednom akcentogenom rečju. Ukoliko ta reč ima silazni akcenat on mora da se pomeri za jedan slog ka početku, odnosno na proklitiku, da se ne bi našao na unutrašnjem slogu izgovorne reči. Prenošenje akcenta na proklitiku naziva se *prokliza* i vrši se po određenim pravilima. Prenose se samo silazni akcenti, dok se uzlazni ne prenose jer mogu ostati na unutrašnjem slogu akcenatske celine. Nakon prenošenja, preneseni akcenat može i dalje biti silazni. Ovakvo prenošenje se naziva *metataksa* i to je tzv. staro prenošenje, predbalkansko (*vȍdu – ù vodu, nȍgu – nà nogu, glâvu – pȍd glāvu*). Ako je proklitika dvosložna, akcenat mora preći na prvi slog proklitike (*ljûde – mȅđu ljūde, lèda – ȉspod leda*). Ukoliko preneseni akcenat menja intonaciju, tj. postaje uzlazni (*metatonija*) u pitanju je novo prenošenje (balkansko) koje se vršilo istovremeno kada i pomeranje silaznih akcenata unutar reči. Kod novog prenošenja akcenat ide na drugi slog dvosložne proklitike jer je uzlazni (*zgràdē – izà zgradē, kȕćē – isprèd kućē, kȍžē – ispòd kožē*). Kada je u pitanju odrična rečca *ne* prokliza je obavezna (*râdim – nè rādīm, mȉslē – nè mislē, znȁjū – nè znajū*).

Kao što se pojedini slogovi u reči razlikuju po prozodijskim obeležjima, tako se i pojedine reči razlikuju u okviru jedne rečenice. Osim leksičkih akcenata, prozodiju rečenice karakterišu i rečenični naglasak i rečenična intonacija. Rečenični naglasak predstavlja naročito isticanje nekog dela rečenice. Brojni eksperimenti su pokazali da upravo rečenična intonacija ima najveći uticaj na kretanje visine glasa u toku rečenice, a da su promene u visini glasa koje unose leksički akcenti lokalnog karaktera (Lehiste & Ivić, 1996).

4.7 SLOG

Slog je najmanja izgovorna jedinica nastala jedinstvenom artikulacijom čija se akustička obeležja svode na talas narastanja i slabljenja sonornosti na osnovu čega se može zaključiti da slog predstavlja maksimum sonornosti ostvaren uz minimum artikulacionog napora (Petrović & Gudurić, 2010). U svom idealnom obliku slog se javlja kao kombinacija šuma i tona, mada u govornom lancu može imati i druge oblike, pri čemu je prisustvo tona obavezno. Ton predstavlja vrhunac sonornosti u segmentu govornog lanca dok se ispred i iza njega raspoređuju glasovi prema pravilima rastuće odnosno opadajuće sonornosti.

Teorijski svaki slog se može razložiti na dva dela: početak i rimu, pri čemu rimu čine jezgro sloga i koda (Petrović & Gudurić, 2010; Mihaljević, 1991). Po engleskoj terminologiji slog se razlaže na *onset* i *rhyme,* pri čemu *nucleus* i *coda* čine *rhyme*. Struktura sloga prikazana je na slici 4.7. Jezgro sloga je nosilac vrhunca sonornosti i predstavlja obavezni deo slogovne strukture. Margine sloga, na uzlaznoj (”uzlaz” ili ”pristup”) i silaznoj padini (”koda” ili ”odstup”), mogu biti različite i od konfiguracije konsonanata ili grupe konsonanata u okviru slogovnog luka zavisiće priroda slogovne strukture.



Slika 4.7 – Struktura sloga

Tipovi sloga u srpskom jeziku mogu biti (Petrović & Gudurić, 2010):

1) Sam vokal V koji se ostvaruje kao jednosložna leksička jedinica (/a/, /u/, /i/, /o/ kao veznici ili predlozi) ili vokal kao deo višesložne leksičke jedinice (/o/-/i/-vi-či-ti, na-/u/-či-ti, /u/-/i/-nat);

2) Grupa CV kao jednosložna leksema (/sa/, /na/, /po/, /ka/, /se/, /je/, /tu/ i sl.) ili elemenat višesložnih reči (/ra/-/do/-/va/-/ti/, /ra/-/di/-/ti/, /pe/-/va/-/ti/);

3) Grupa VC kao jednosložna leksema (/od/, /iz/, /uz/) ili kao deo višesložne leksičke jedinice u inicijalnom slogu u slučaju deljive suglasničke grupe, pri čemu se pod deljivom suglasničkom grupom podrazumevaju dva sukcesivno izgovorena suglasnika gde se artikulacija prvog svodi na imploziju, a artikulacija drugog na eksploziju (/ob/-da-ri-ti, /op/-ko-li-ti, /ot/-klo-ni-ti)

4) Grupa CVC u jednosložnim rečima (/kad/, /sad/, /rad/, /dar/, /nad/, /kod/ i sl.) ili u slučaju deljive suglasničke grupe na početku, unutar ili na kraju reči (/kat/-/kad/, /nad/-gle-da-ti, /pod/-ve-sti);

5) Grupa CCV(C) (/sta/-rac, /bra/-ća, /mla/-do-/sti); kao i u jednosložnim rečima (/grad/, /mlad/, /prav/, /snop/, /stan/) ili u višesložnim rečima tipa (/pred/-gra-đe, nad-/grad/-nja i sl.);

6) Grupa CCCV(C) (/sple(t)/, /stra(h)/, /spra(t)/, /skra/-ti-ti, /zgra/-bi-ti, /smla/-či-ti; gde bi se mogla dodati i retka jenosložna reč sa grupom CCCV(CC) (/stra(st)/, /štra(nd)/, ali i grupa CCCV(CCC) u pozajmljenici (/brja(nsk)/);

7) Grupa CCCCV u primerima tipa bra-/tstvo/, poku-/ćstvo/, stude-/ntski/, ke-/ltski/, go-/lfski/, ža-nda-/rmski/ nikada se u jeziku ne može javiti u inicijalnom položaju u reči jer četiri neslogotvorna elementa predstavljaju veliko opterećenje u funkcionisanju artikulacionih mehanizama. Ova suglasnička grupa predvajaće se u sredini reči u zavisnosti od individualne artikulacije govornika. Razlika u predvajanju na slogove može se izvesti na osnovu akustičko-artikulacione povezanosti glasova u glasovnom nizu, pa se granica sloga može naći između suglasnika, pri čemu će se na jednoj strani pojaviti slog zatvoren suglasnikom ili suglasničkom grupom, a na drugoj slog koji počinje suglasničkom grupom koja se može naći i u inicijalnom delu reči: bra*t-stv*o, po-ku*ć-stv*o, stu-de*nt-sk*i, ke*lt-sk*i, go*lf-sk*i, žan-da*rm-sk*i.

Najčešća struktura sloga u srpskom jeziku, kao i u većini indoevropskih jezika, jeste struktura CV i ona čini 60.5% svih slogovnih struktura. Slogovi sa strukturom CCV zastupljeni su u približno 11.5% slogova, oko 10.87% su slogovi sa strukturom CVC, dok slogovi sa strukturom V učestvuju sa 9.7% u glasovnim nizovima srpskog jezika. Sve ostale kombinacije zajedno zastupljene su u nešto manje od 8% slučajeva (Jovičić, 1999). Otvoren slog, odnosno tip sloga koji se završava vokalom je u srpskom jeziku najzastupljeniji tip sloga i čini oko 87.5% svih slogovnih realizacija.

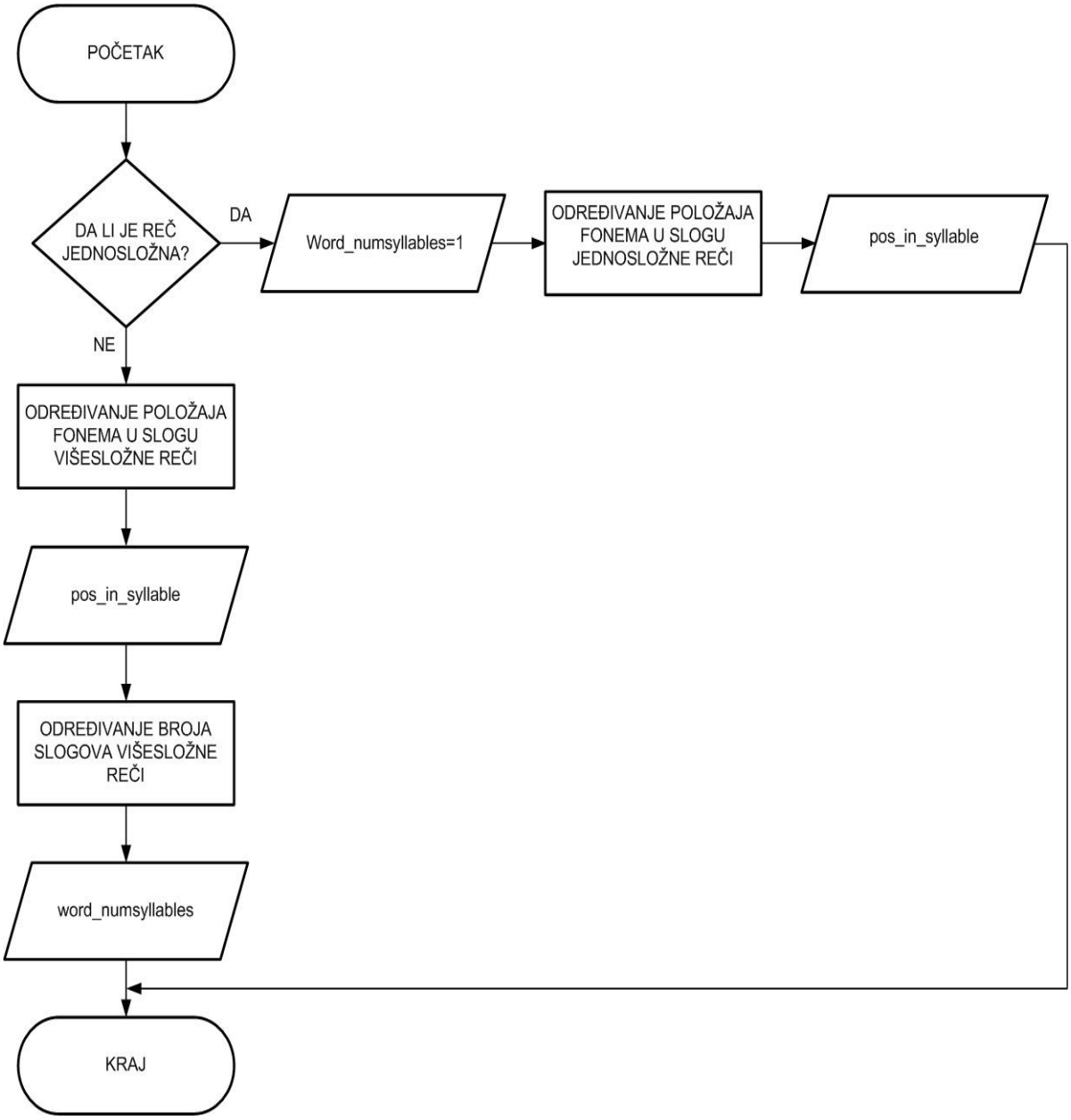
4.7.1 Podela na slogove

U većini jezika podela reči na slogove vrši se prema unapred utvrđenim manje ili više strogim pravilima koja u različitim jezicima mogu biti različita. Stoga, grupa koja se u jednom jeziku predvaja na dva sloga, u drugom jeziku se smatra nedeljivom. Ovakva shvatanja, imaju s

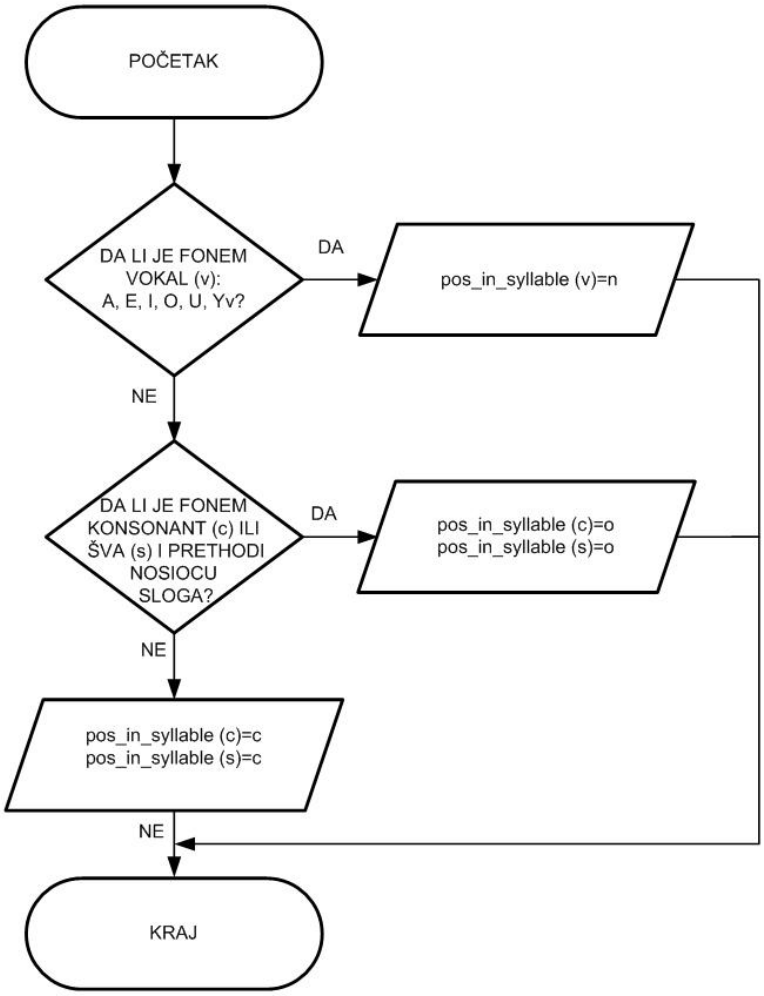
jedne strane uporište u akustičko-artikulacionim osobenostima generisanja glasovnih nizova, a sa druge u perceptivnim sposobnostima pojedinca (Petrović & Gudurić, 2010).

Takođe, segmentacija na slogove se u pojedinim slučajevima vrši jednoobrazno, dok se u nekim drugim slučajevima javlja više mogućnosti u zavisnosti od konkretne artikulacije glasovnog niza.

U okviru ove doktorske disertacije razvijen je algoritam za podelu reči na slogove u srpskom jeziku. Pomenuti algoritam, kao i njegovi blokovi za određivanje položaja fonema u slogu jednosložne i višesložne reči šematski su prikazani na slikama 4.7, 4.8 i 4.9.



Slika 4.8 – Struktura algoritma za podelu reči na slogove u srpskom jeziku

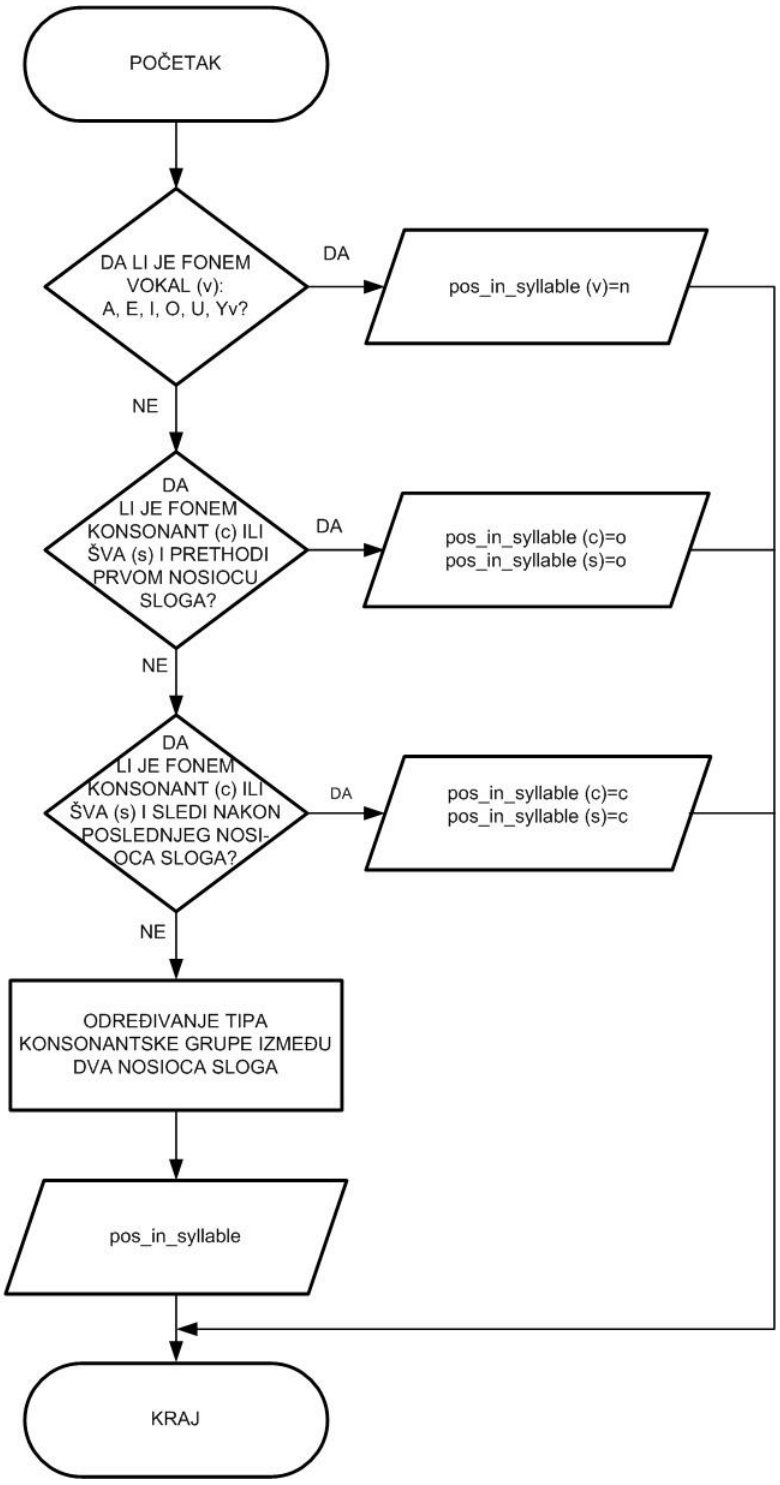


Slika 4.9 – Određivanje položaja fonema u slogu jednosložne reči

Određivanje tipa konsonantske grupe između dva nosioca sloga

Konsonanti između dva nosioca sloga mogu biti:

1. jedan intervokalski konsonant
2. dvočlana konsonantska grupa
3. tročlana konsonantska grupa
4. četvoročlana konsonantska grupa
5. petočlana konsonantska grupa
6. nema konsonanata između dva nosioca sloga



Slika 4.10 – Određivanje položaja fonema u slogu višesložne reči

1. intervokalski konsonant

Konsonant između dva nosioca sloga (n) pripada slogu koji formira naredni nosilac sloga i taj konsonant predstavlja onset narednog nosioca sloga, odnosno pos_in_syllable=o

2. dvočlana konsonantska grupa

Konsonantska grupa koju formiraju dva uzastopna konsonanta može biti deljiva ili nedeljiva.

a) deljiva grupa

Ako je konsonantska grupa deljiva tada prvi konsonant predstavlja kodu prethodnog nosioca sloga, a drugi konsonant predstavlja onset narednog nosioca sloga.

K1K2:K1+K2

K1: pos_in_syllable=c

K2: pos_in_syllable=o

b) nedeljiva grupa

Ako je konsonantska grupa nedeljiva tada oba konsonanta predstavljaju onset narednog nosioca sloga.

K1K2

Nedeljive konsonantske grupe su (Petrović & Gudurić, 2010):

1. /cv/, /šp/, /št/, /ht/, /žv/, /šv/, /sv/, /hv/, /žg/
2. /pr/, /pl/, /tr/, /kr/, /kl/, /klj/, /knj/, /čr/, /čl/, /šr/, /šlj/, /šm/, /šn/, /sr/, /slj/, /cr/, /fr/, /fl/, /flj/, /hr/, /hl/, /vr/, /vl/, /gr/, /gl/, /glj/, /gm/, /gn/, /gnj/, /žr/
3. /mr/, /ml/

3. tročlana konsonantska grupa

Tročlanu konsonantsku grupu čine tri uzastopna konsonanta. Ona može biti podeljena na dva načina (Petrović & Gudurić, 2010).

Ukoliko je u pitanju neka od sledećih kombinacija tri uzastopna konsonanta:

1. **K1**: sonant; **K2**: frikativ; **K3**: opstruent
2. **K1**: opstruent; **K2**: /m/; **K3**:/l/,/n/, /r/, /nj/
3. **K1**: opstruent; **K2**: opstruent; **K3**: sonant
4. **K1**:opstruent; **K2**: /s/, /š/; **K3**: /k/, /t/
5. **K1**: /z/, /d/, /b/, /s/, /t/, /p/, /j/; **K2**:opstruent; **K3**:opstruent

tročlana konsonantska grupa deli se kao pod a) **K1K2K3: K1+K2K3** u protivnom deli se kao pod b).

a) K1K2K3: K1+K2K3

K1 predstavlja kodu prethodnog nosioca sloga
K1: pos_in_syllable=c

K2K3 predstavljaju onset narednog nosioca sloga
K2, K3: pos_in_syllable=o

b) **K1K2K3**: **K1K2**+**K3**

K1K2 predstavljaju kodu prethodnog nosioca sloga
K1, K2: pos_in_syllable=c

K3 predstavlja onset narednog nosioca sloga
K3: pos_in_syllable=o

4. četvoročlana konsonantska grupa

Četvoročlanu konsonantsku grupu čine četiri uzastopna konsonanta. Ona se u srpskom jeziku najčešće deli na dva načina (Petrović & Gudurić, 2010).

Ukoliko je u pitanju neka od sledećih kombinacija četiri uzastopna konsonanta:

1. **K1**: konsonant; **K2**: /s/, **K3**: /t/, **K4**:/v/
2. **K1**: /d/, /t/, /j/, **K2**: frikativ, **K3**: ploziv, **K4**: sonant

četvoročlana konsonantska grupa deli se kao pod a) **K1K2K3K4: K1+K2K3K4** u protivnom deli se kao pod b).

a) K1K2K3K4: K1+K2K3K4

K1 predstavlja kodu prethodnog nosioca sloga
K1: pos_in_syllable=c

K2K3K4 predstavljaju onset narednog nosioca sloga
K2, K3, K4: pos_in_syllable=o

b) K1K2K3K4: **K1K2**+**K3K4**

K1K2 predstavljaju kodu prethodnog nosioca sloga
K1, K2: pos_in_syllable=c

K3K4 predstavljaju onset narednog nosioca sloga
K3, K4: pos_in_syllable=o

5. petočlana konsonantska grupa

Petočlanu konsonantsku grupu čini pet uzastopnih konsonanta. Ona se u najvećem broju slučajeva u srpskom jeziku deli (Petrović & Gudurić, 2010):

K1K2K3K4K5: **K1K2**+**K3K4K5**

K1K2 predstavljaju kodu prethodnog nosioca sloga

K1, K2: pos_in_syllable=c

K3K4K5 predstavljaju onset narednog nosioca sloga

K3, K4, K5: pos_in_syllable=o

6. nema konsonanata između dva nosioca sloga

Između dva nosioca sloga (vokala) nema nijednog konsonanta tada su u pitanju dva različita sloga ili ta dva vokala predstavljaju diftong i čine jednog nosioca sloga, pri čemu prvi vokal ima pos_in_syllable=n1 a drugi pos_in_syllable=n2

U nastavku su navedeni konsonanti, kao i odgovarajuće grupe kojima pripadaju.

Konsonanti: /b/, /v/, /g/, /d/, /đ/, /ž/, /z/, /j/, /k/, /l/, /lj/, /m/, /n/, /nj/, /p/, /r/, /s/, /t/, /ć/, /f/, /h/, /c/, /č/, /dž/, /š/

Opstruenti: /b/, /v/, /g/, /d/, /đ/, /ž/, /z/, /j/, /k/, /p/, /s/, /t/, /ć/, /f/, /h/, /c/, /č/, /dž/, /š/

Sonanti: /l/, /lj/, /r/, /m/, /n/, /nj/

Plozivi: /p/, /t/, /k/, /b/, /d/, /g/

Frikativi: /s/, /š/, /f/, /h/, /z/, /ž/, /v/

Afrikati: /c/, /ć/, /č/, /đ/, /dž/

Kod višesložne reči broj slogova u reči jednak je zbiru ukupnog broja nosilaca sloga (n) i ukupnog broja diftonga (n1), odnosno word_numsyllables=n+n1.

5. GOVORNA BAZA I FAKTORI KORIŠĆENI U PROCESU MODELOVANJA TRAJANJA GLASOVA U SRPSKOM JEZIKU

5.1 GOVORNA BAZA

Govorna baza korišćena prilikom modelovanja trajanja glasova u srpskom jeziku sadrži približno 2000 rečenica i oko 16000 reči. Snimanje govorne baze obavljeno je u studiju, pri čemu je korišćen glas profesionalne radijske spikerke koja govori standardni ekavski dijalekt. Prilikom snimanja govornica je bila instruisana da govori posebno razgovetno, kao i da tekst izgovara nepromenjenom jačinom glasa, ni suviše brzo ni suviše sporo, ujednačenim tonom, bez emocija uz prirodnu i jasnu artikulaciju glasova. Kod baza koje obuhvataju kontinualan govor u izgovoru se zahteva što prirodnija intonacija jer se očekuje da će pri sintezi biti spajani i duži segmenti, te je poželjno očuvati njihova prirodna prozodijska obeležja. Najveći deo sadržaja govornog korpusa predstavljaju tekstovi iz dnevne štampe, pisani publicističkim stilom, koji se tipično koriste za ovakve namene. Nakon snimanja govorne baze izvršena je njena fonetska i prozodijska anotacija.

Fonetska anotacija govorne baze, odnosno njeno labeliranje podrazumeva postavljanje granica između jedinica koje pripadaju unapred definisanom skupu jedinica kao što su glasovi. U suštini, ono se svodi na smeštanje informacija o tim jedinicama, kao što su početni i završni trenuci, u posebnu bazu podataka.

Labeliranje govorne baze izvršeno je na nivou fonema, s tim što je za pojedine klase fonema izvršeno labeliranje i karakterističnih supfonemskih jedinica. Primera radi, plozivi i afrikati bili su obeležavani kao parovi polufonema. Ove parove u slučaju ploziva čine okluzija i eksplozija, a u slučaju afrikata okluzija i frikcija. Treba svakako napomenuti da određivanje granica između fonema nije trivijalan zadatak. Poznato je da fonemi imaju svoja artikulaciono-akustička svojstva koja razlikuju jedan fonem od drugog. Međutim, fonemi se po pravilu ne javljaju izolovano već u okruženju drugih fonema, što dovodi do promene tih svojstava i pojave poznate pod nazivom *koartikulacija*. U artikulacionom smislu koartikulacija je posledica činjenice da govorni aparat mora da pređe iz položaja karakterističnog za artikulaciju jednog fonema u položaj koji odgovara artikulaciji drugog. U akustičkom smislu koartikulacija se manifestuje kao promena položaja formanata u spektru vokala, odnosno kao promena položaja koncentrata energije u spektru konsonanata. Zadatak labeliranja je otežan jer se s obzirom na

koartikulaciju susednih fonema ne može uočiti oštar prelaz sa jednog fonema na drugi. Svi slučajevi pojave oštećenih ili delimično oštećenih glasova takođe su bili evidentirani, zajedno sa podatkom o stepenu oštećenja. Ovo je urađeno da bi se izbeglo njihovo korišćenje prilikom sinteze u kontekstu u kom su tipično dobro artikulisani, kao na primer u okviru naglašenog sloga a takođe i da bi takvi glasovi bili izuzeti u procesu modelovanja trajanja. Labeliranje je izvršeno automatski korišćenjem AlfaNumASR sistema za prepoznavanje kontinualnog govora (Delić et al., 2010). Provera i korekcija dobijenih rezultata izvršena je ručno, na osnovu talasnog oblika signala, njegovog spektrograma i auditorne percepcije. Pri ručnoj proveri i korekciji korišćen je softveriski alat AlfaNum TTSLabel (Delić et al., 2010).

Prozodijska anotacija govorne baze obuhvatila je označavanje leksičkog akcenta, rečeničnog fokusa i različitih perceptivno uočenih nivoa pauze. Prilikom označavanja leksičkog akcenta naglašenom vokalu dodeljena je jedna od četiri moguće vrste akcenta u srpskom jeziku ili posleakcenatska dužina. Kod označavanja rečeničnog fokusa pojedinim rečima dodeljen je pozitivan fokus ukoliko je u pitanju naročito istaknuta reč, odnosno negativan fokus ako se radi o relativno nebitnoj reči. Prozodijska anotacija rađena je ručno pomoću softverskog alata AlfaNum TTSLabel (Delić et al., 2010).

Pored navedenih informacija za svaku reč u govornoj bazi obeležena je vrsta reči, kao i odgovarajuća akcenatska konfiguracija.

5.2 FAKTORI KOJI UTIČU NA TRAJANJE GLASOVA U SRPSKOM JEZIKU

U procesu modelovanja trajanja neophodna komponenta TTS sistema, koja prethodi modulu za određivanje trajanja određenog govornog segmenta u datom kontekstu, jeste modul za automatsko generisanje odgovarajućeg vektora obeležja kojim se predstavlja svaki fonem u govornoj bazi. Elementi vektora obeležja opisuju određeni govorni segment i kontekst u kome se on nalazi, pri čemu je vrednost svakog obeležja zapravo jedan od mogućih nivoa faktora koji utiče na trajanje govornog segmenta.

Ako je određeni govorni segment u bazi predstavljen preko odgovarajućeg vektora obeležja f i ako faktor f_j ukazuje na prisustvo leksičkog akcenta sloga i uzima međusobno isključive vrednosti iz skupa {naglašen, nenaglašen}, tada element vektora obeležja ima jednu od mogućih vrednosti faktora f_j. Prostor proizvoda svih faktora $f_1 \times f_2 \times \ldots \times f_n$ naziva se prostor obeležja F. Međutim, zbog različitih fonoloških i drugih lingvističkih ograničenja, nisu sve kombinacije različitih vrednosti faktora dozvoljene u nekom jeziku. Stoga, lingvistički prostor

koji definišu samo oni vektori obeležja koji se zaista pojavljuju u određenom jeziku značajno je manji od prostora obeležja i predstavlja podskup prostora obeležja. S druge strane, broj lingvistički mogućih kombinacija različitih vrednosti faktora u nekom jeziku je izuzetno velik i snimanje takvog govornog korpusa prevazilazi razumne vremenske okvire (van Santen, 1992). Shodno tome, prilikom izbora materijala za snimanje govorne baze velika pažnja usmerena je ka pronalaženju materijala koji će sadržati što je moguće veći broj različitih lingvistički mogućih kombinacija u cilju ostvarenja što veće pokrivenosti lingvističkog prostora, mada i pored svih napora govorni korpus često predstavlja samo mali podskup lingvističkog prostora (van Santen, 1994).

Na osnovu najuticajnijih faktora koje su autori koristili prilikom modelovanja trajanja govornih segmenata u različitim jezicima (Klatt, 1976; van Santen, 1994; Campbell, 1992; Moebius & van Santen, 1996; Venditti & van Santen, 1998; Febrer et al., 1998; Batušek, 2002; Lazaridis et al., 2007), kao i na osnovu rezultata dosadašnjih istraživanja koja se odnose na uticaj različitih faktora na trajanje fonema u srpskom jeziku (Sovilj-Nikić, 2007; Lehiste & Ivić, 1996; Marković & Milićev, 2009) izabrani su faktori koji će u nastavku istraživanja biti uzeti u obzir prilikom daljeg razvoja modela trajanja glasova u sintezi govora na srpskom jeziku. Svi faktori, koji će kasnije biti navedeni, ekstrahovani su iz govorne baze na srpskom jeziku snimljene za potrebe postojećeg sintetizatora govora (Sečujski et al., 2007) i mnogobrojnih istraživanja koja se sprovode u cilju njegovog unapređenja.

Svaki fonem u govornoj bazi predstavljen je preko odgovarajućeg vektora obeležja koji opisuje dati govorni segment i kontekst u kom se taj fonem nalazi. U nastavku su navedeni faktori koji su korišćeni u procesu modelovanja trajanja glasova u okviru ove disertacije, kao i njihove moguće vrednosti u srpskom jeziku. Faktori su razvrstani prema domenima njihovog delovanja.

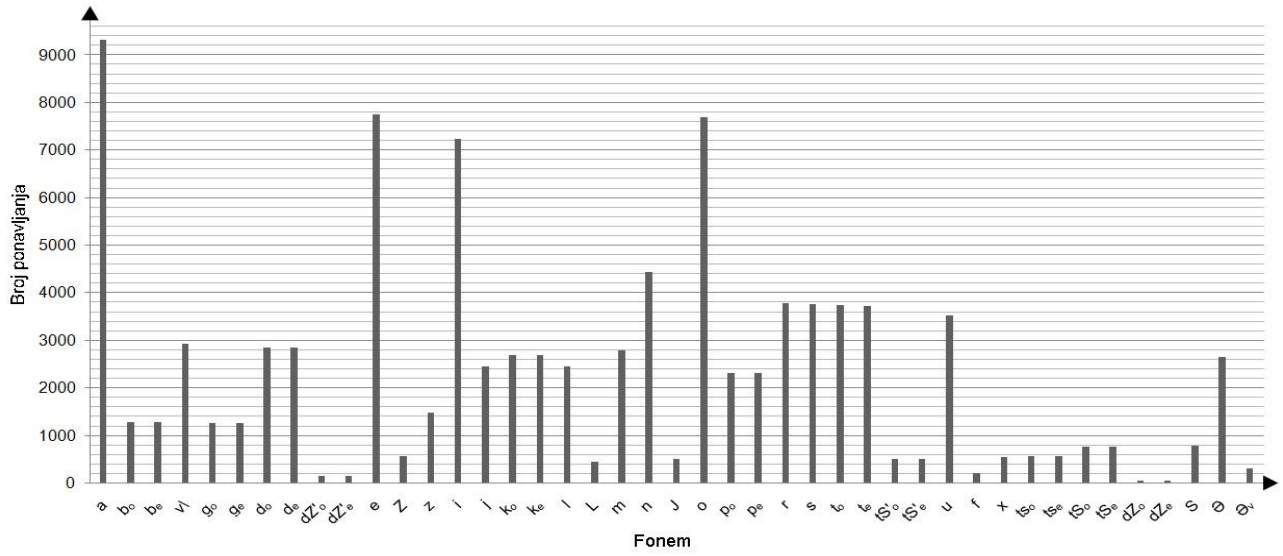
- **Trenutni segment**

identitet segmenta: Može imati jednu od 43 različite vrednosti u koje spadaju pet vokala srpskog jezika, dve različite realizacije poluglasa (šva) /ə/ i 25 konsonanata. S obzirom na to da su plozivi i afrikati u bazi obeležavani kao parovi polufonema, koje u slučaju ploziva čine okluzija i eksplozija a u slučaju afrikata okluzija i frikcija, ukupan broj različitih konsonanata je u ovom slučaju 36. Napravljena je distinkcija između dve varijante poluglasa /ə/ koji se u govoru javlja u situacijama kada se suglasnik /r/ nađe u suglasničkom okruženju, odnosno u vokalskoj upotrebi (Gudurić & Petrović, 2005). Kod vokalske upotrebe, glas /r/ može biti nosilac sloga i upravo takvu realizaciju vokalnog elementa /ə/ treba razlikovati od one u kojoj /r/ nije slogotvorno.

U tabeli 5.1 dati su SAMPA (engl. *Speech Assesment Method for Phonetic Alphabet*) simboli za svaki od fonema korišćenih u procesu modelovanja trajanja a na slici 5.1 prikazan je broj pojavljivanja svakog fonema u govornoj bazi.

Fonem	SAMPA simbol	Fonem	SAMPA simbol
/a/	a	/o/	o
/b/$_o$	b$_o$	/p/$_o$	p$_o$
/b/$_e$	b$_e$	/p/$_e$	p$_e$
/v/	v\	/r/	r
/g/$_o$	g$_o$	/s/	s
/g/$_e$	g$_e$	/t/$_o$	t$_o$
/d/$_o$	d$_o$	/t/$_e$	t$_e$
/d/$_e$	d$_e$	/ć/$_o$	tS'$_o$
/dj/$_o$	dZ'$_o$	/ć/$_e$	tS'$_e$
/dj/$_e$	dZ'$_e$	/u/	u
/e/	e	/f/	f
/ž/	Z	/h/	x
/z/	z	/c/$_o$	ts$_o$
/i/	i	/c/$_e$	ts$_e$
/j/	j	/č/$_o$	tS$_o$
/k/$_o$	k$_o$	/č/$_e$	tS$_e$
/k/$_e$	k$_e$	/dž/$_o$	dZ$_o$
/l/	l	/dž/$_e$	dZ$_e$
/lj/	L	/š/	S
/m/	m	/ə/	ə
/n/	n	/ə/$_v$	ə$_v$
/nj/	J		

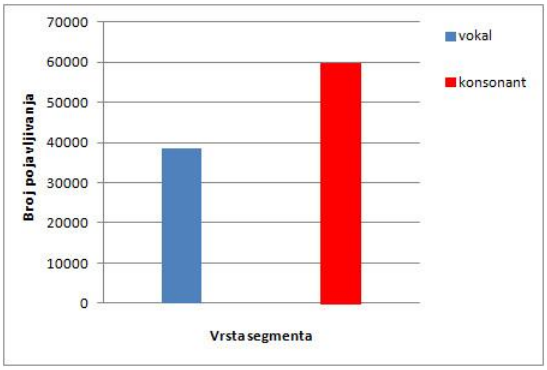
Tabela 5.1 – SAMPA simboli fonema srpskog jezika



Slika 5.1 – Raspodela fonema srpskog jezika u govornoj bazi

vrsta segmenta: vokal, konsonant

Ukupan broj vokala i konsonanata u govornoj bazi predstavljen je na slici 5.2.

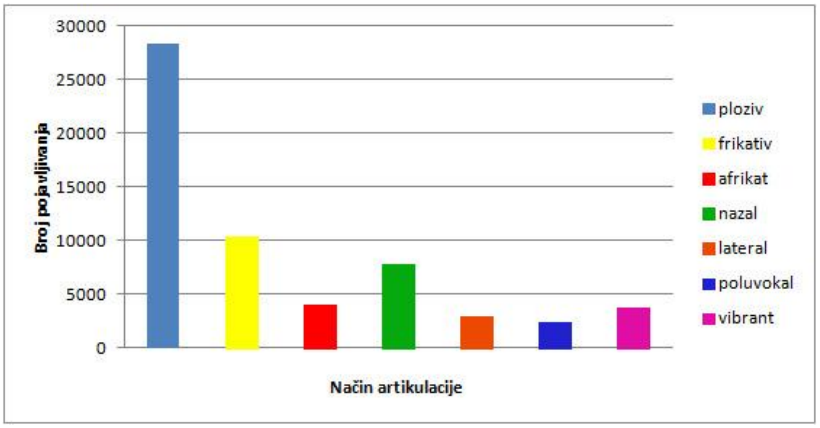


Slika 5.2 – Raspodela vokala i konsonanata u govornoj bazi

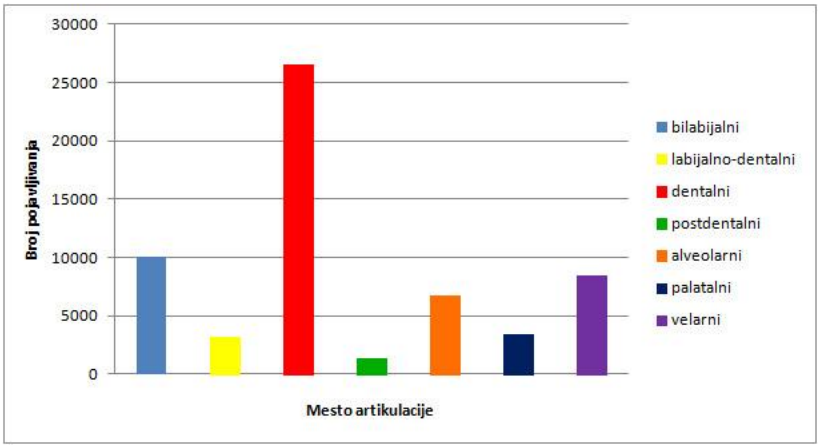
način artikulacije (za konsonante): ploziv, frikativ, afrikat, nazal, lateral, poluvokal, vibrant

mesto artikulacije (za konsonante): bilabijalni, labijalno-dentalni, dentalni, postdentalni, alveolarni, palatalni, velarni

Raspodela konsonanata u govornoj bazi prema načinu, odnosno mestu artikulacije prikazana je na slikama 5.3 i 5.4.



Slika 5.3 – Raspodela konsonanata prema načinu artikulacije



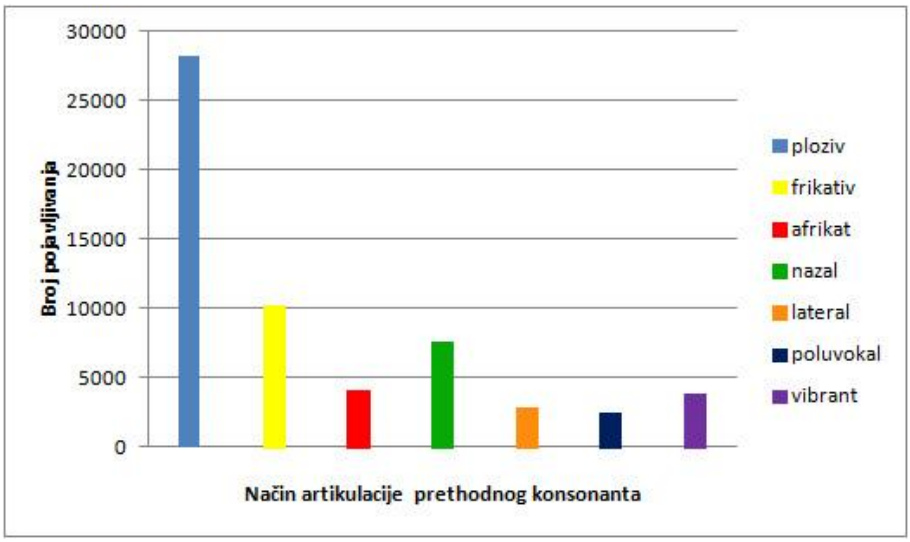
Slika 5.4 – Raspodela konsonanata prema mestu artikulacije

- **Neposredno okruženje (prethodni i naredni segment)**

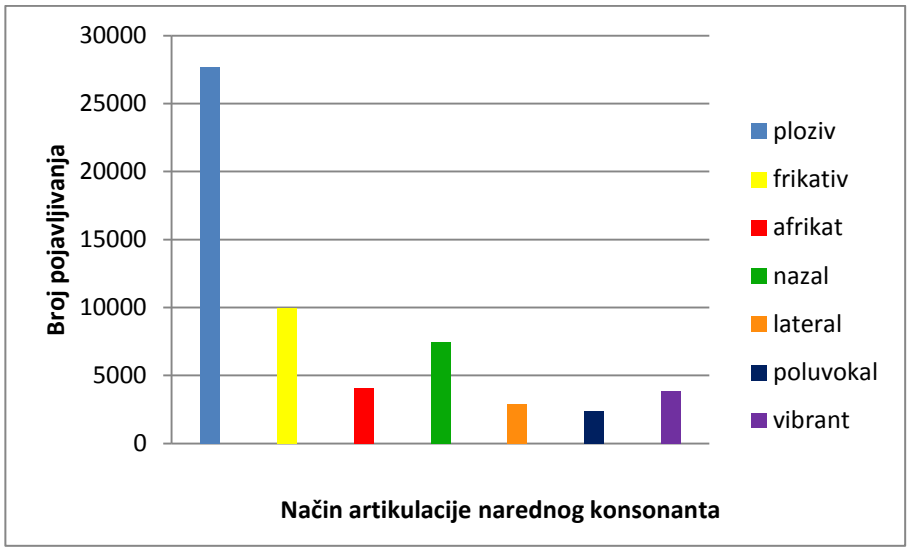
vrsta segmenta: vokal, konsonant, tišina

način artikulacije (za konsonante): ploziv, frikativ, afrikat, nazal, lateral, poluvokal, vibrant

Na slici 5.5 prikazana je raspodela prema načinu artikulacije konsonanata koji se nalaze u neposrednom okruženju trenutnog segmenta.



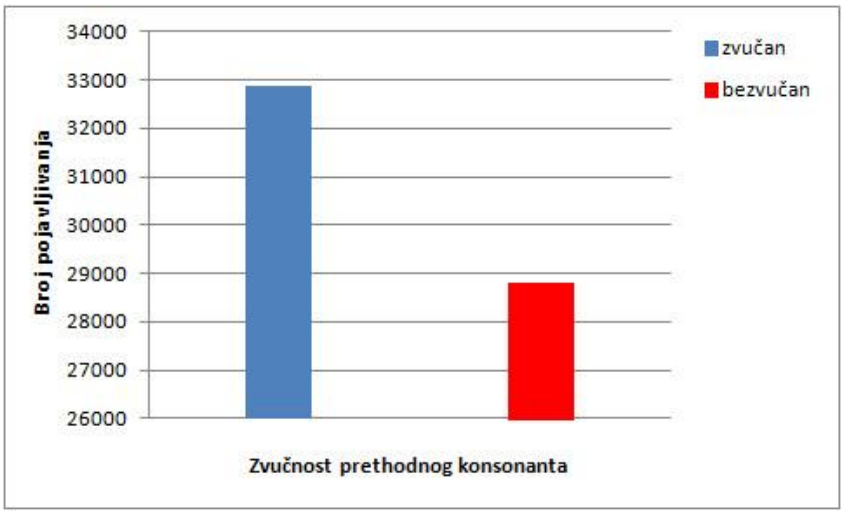
Slika 5.5 a – Raspodela prethodnog konsonanta prema načinu artikulacije



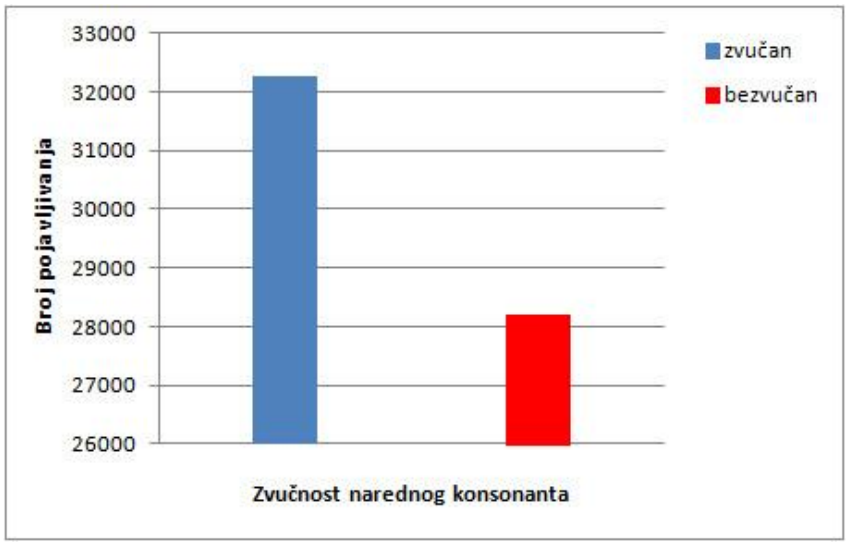
Slika 5.5 b – Raspodela narednog konsonanta prema načinu artikulacije

zvučnost: zvučan, bezvučan

Ukupan broj zvučnih, odnosno bezvučnih fonema u neposrednom okruženju datog segmenta prikazan je na slici 5.6.



Slika 5.6 a – Raspodela prethodnog konsonanta prema zvučnosti



Slika 5.6 b – Raspodela narednog konsonanta prema zvučnosti

U mnogim istraživanjima pokazano je da zvučnost naročito narednog konsonanta u velikoj meri utiče na trajanje fonema, odnosno doprinosi njegovom produženju (Klatt, 1976; Sovilj-Nikić, 2007; Campbell, 1992; Öztürk, 2005). Takođe je ustanovljeno da bezvučni konsonanti traju duže od zvučnih (Öztürk, 2005).

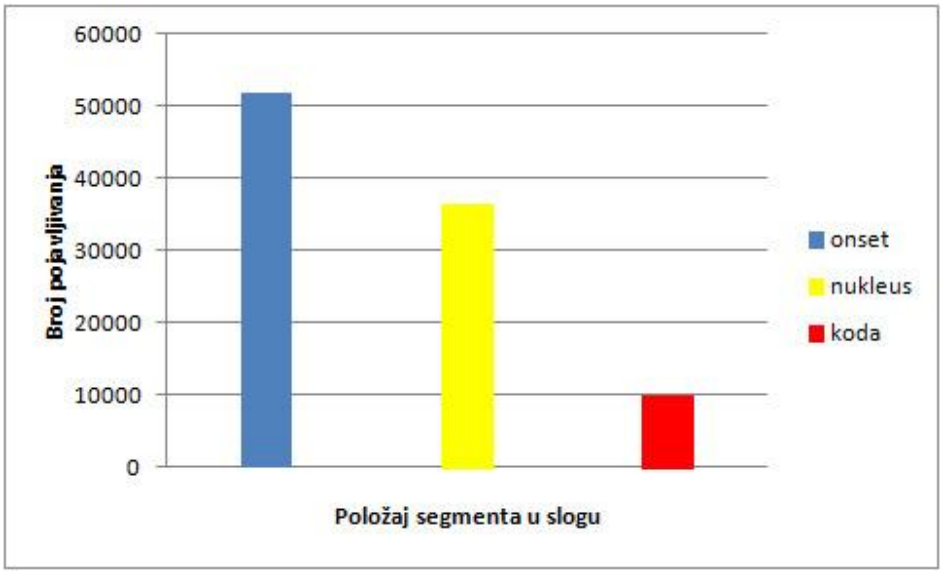
Mesto artikulacije prethodnog i narednog konsonanta se ne uzima u obzir jer su prethodna istraživanja pokazala da ne predstavlja relevantan faktor (Crystal & House, 1988).

- **Položaj segmenta u slogu**

početni položaj: da, ne

položaj u slogu: onset, jezgro (nukleus), koda

Raspodela fonema prema položaju u slogu prikazana je na slici 5.7.

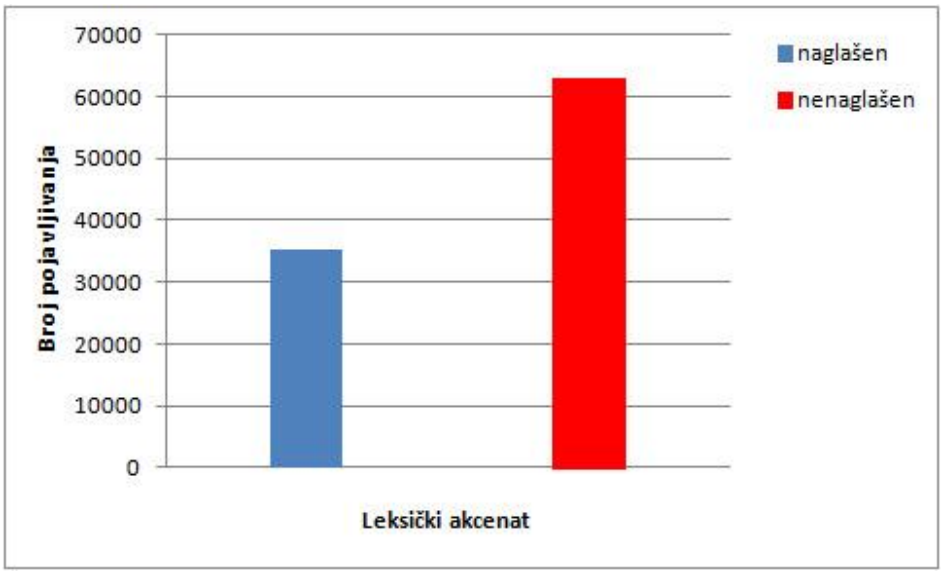


Slika 5.7 – Raspodela fonema prema položaju u slogu

- **Slog**

leksički akcenat: naglašen, nenaglašen

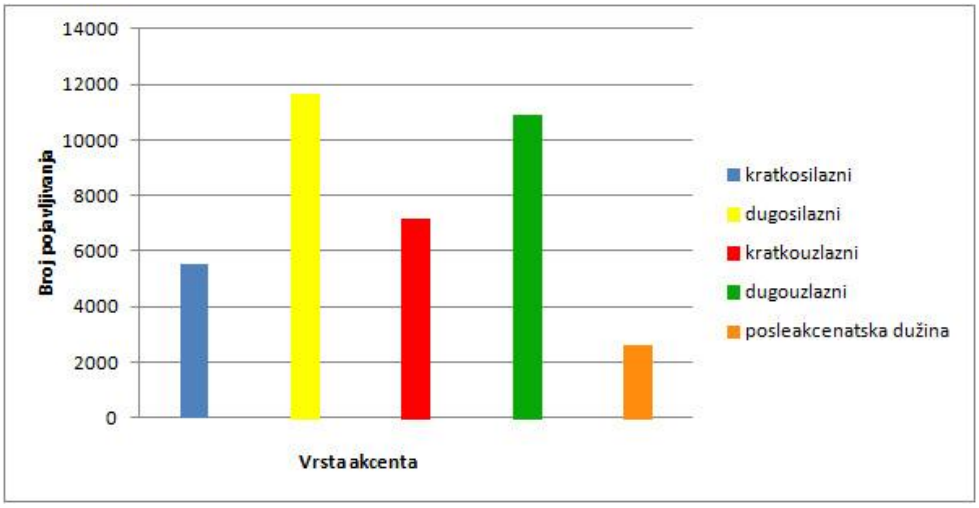
Slika 5.8 prikazuje ukupan broj fonema koji se u govornoj bazi nalaze u naglašenom, odnosno nenaglašenom slogu.



Slika 5.8 – Raspodela leksičkog akcenta u govornoj bazi

vrsta akcenta: kratkosilazni, dugosilazni, kratkouzlazni, dugouzlazni, posleakcenatska dužina

Raspodela fonema u govornoj bazi u slogovima sa jednom od četiri vrste akcenta ili posleakcenatskom dužinom prikazana je na slici 5.9.



Slika 5.9 – Raspodela vrste akcenta u govornoj bazi

- **Položaj sloga u reči**

početni: da, ne

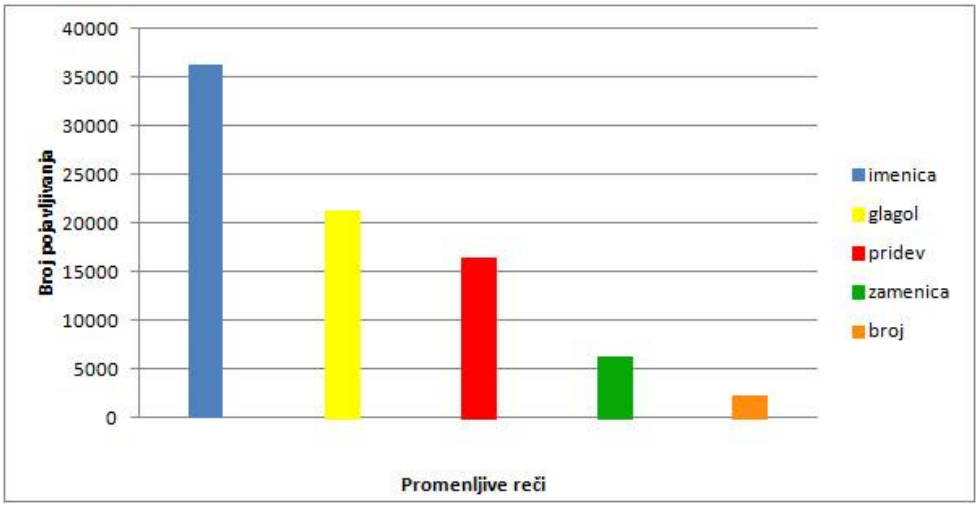
krajnji: da, ne

Smatra se da produženje trajanja ne samo vokala već i konsonanata na kraju reči ima perceptivnu funkciju, odnosno da doprinosi označavanju granica između reči (Bakran, 1996). Stoga, ovaj efekat treba uzeti u obzir u sintezi govora jer on ne doprinosi samo prirodnosti govora, nego i njegovoj razumljivosti.

- **Reč**

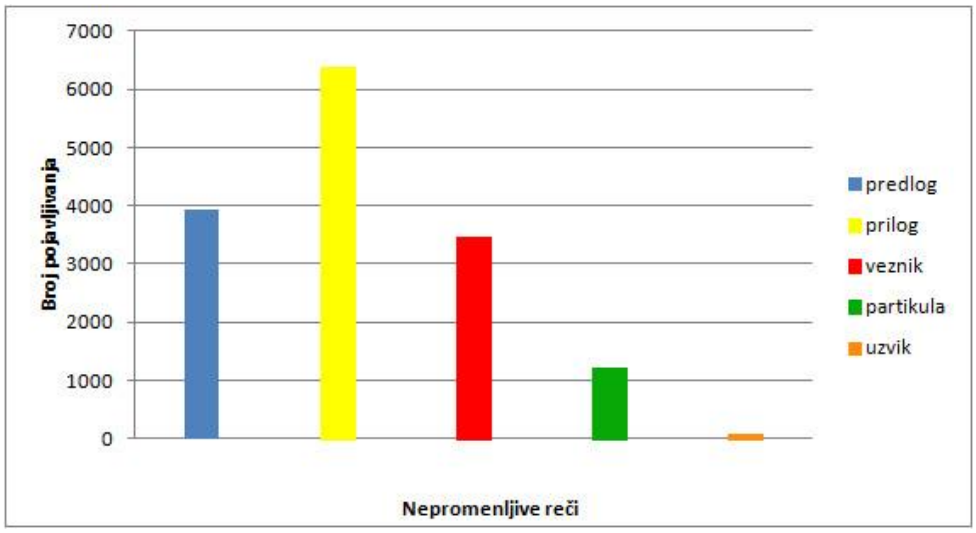
vrsta reči: promenljive reči: imenica, glagol, pridev, zamenica, broj

nepromenljive reči: predlog, prilog, veznik, partikula, uzvik



Slika 5.10 – Raspodela fonema u promenljivim rečima

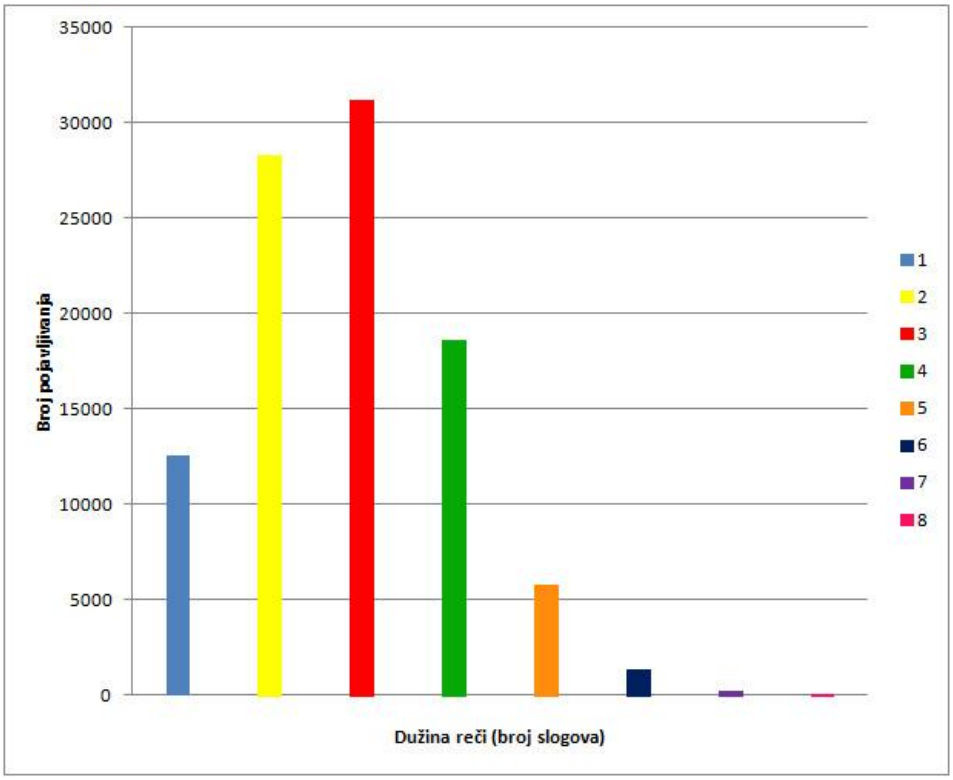
Na slici 5.10 prikazana je raspodela fonema koji se nalaze u promenljivim rečima u govornoj bazi, dok slika 5.11 prikazuje raspodelu fonema u nepromenljivim rečima.



Slika 5.11 – Raspodela fonema u nepromenljivim rečima

dužina reči: broj slogova u reči

Na slici 5.12 dat je histogram koji prikazuje raspodelu fonema u govornoj bazi u rečima različite dužine koja je izražena brojem slogova u reči.



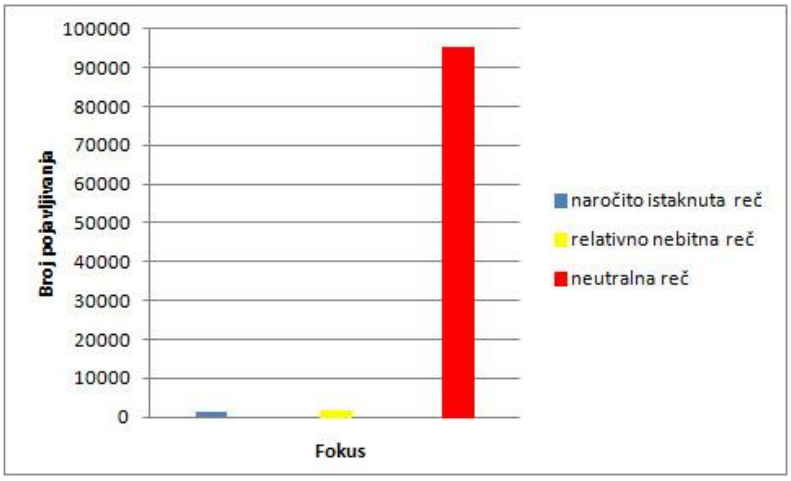
Slika 5.12 – Raspodela fonema u rečima različite dužine

U prethodnim istraživanjima pokazano je da je trajanje govornih segmenata obrnuto proporcionalno broju slogova u reči, odnosno da povećanje broja slogova u reči doprinosi skraćenju trajanja govornih segmenata (Sovilj-Nikić, 2007; Marković & Milićev, 2009).

- **Fokus**

fokus: naročito istaknuta reč, relativno nebitna reč, neutralna reč

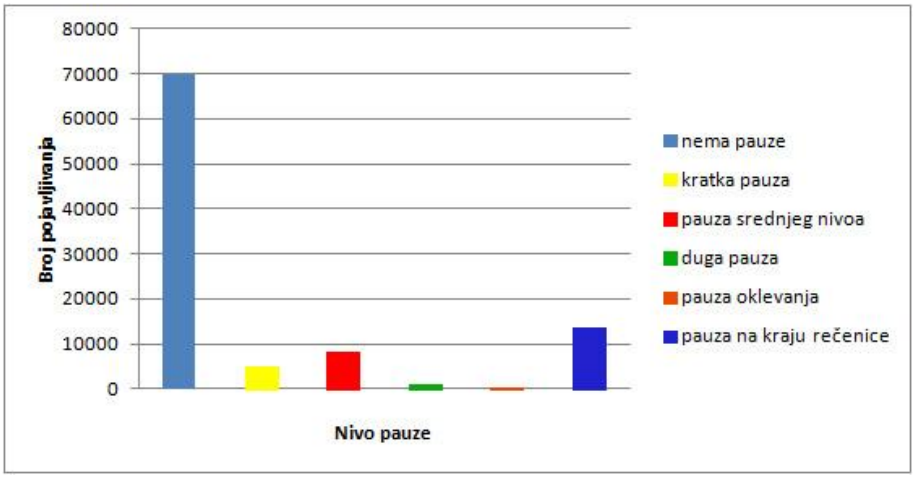
Histogram na slici 5.13 prikazuje raspodelu fonema u govornoj bazi u rečima različitog rečeničnog fokusa, pri čemu je napravljena distinkcija između pozitivnog fokusa, negativnog fokusa i neutralne reči.



Slika 5.13 – Raspodela fonema u rečima različitog fokusa

- **Položaj reči u frazi**

nivo pauze: nema pauze, kratka pauza, pauza srednjeg nivoa, duga pauza, pauza oklevanja, pauza na kraju fraze.



Slika 5.14 – Raspodela fonema u rečima različitih nivoa pauze

Na slici 5.14 histogramom je prikazana raspodela fonema u govornom korpusu u rečima različitih nivoa pauze. Različiti nivoi pauze odgovaraju različitim perceptivno uočenim pauzama, pri čemu ukoliko se pauza poklapa sa intervalom tišine moguće je napraviti razliku između inicijalnog, medijalnog i finalnog položaja reči u frazi. Efekat produženja trajanja segmenata ispred pauze primećen je u mnogim jezicima i smatra se univerzalnom pojavom (van Santen, 1992; van Santen, 1993; Bakran, 1996; Öztürk, 2005).

6. METODE AUTOMATSKOG UČENJA KORIŠĆENE U PROCESU MODELOVANJA TRAJANJA GLASOVA

Prilikom modelovanja trajanja glasova u okviru ove disertacije korišćeno je više različitih metoda automatskog učenja (engl. *machine learning*) u koje se ubrajaju stabla odluke (engl. *decision trees*), linearna regresija, kao i nekoliko meta algoritama kao što su aditivna regresija, *bagging* i *stacking*. Pomenute metode će u nastavku biti ukratko prikazane.

6.1 STABLA ODLUKE

CART (engl. *Classification and Regression Trees*) metodu razvio je 1984. godine Leo Breiman (Breiman, 1984). Ova tehnika razvijena kao spoj statistike i veštačke inteligencije poseduje niz prednosti i kao takva danas predstavlja jednu od najčešće primenjivanih metoda za modelovanje trajanja govornih segmenata. Jedna od osnovnih prednosti CART algoritma jeste mogućnost validacije razvijenog modela, što se u praksi najčešće vrši procenom performansi modela na podacima koji nisu korišćeni u fazi obuke. Takođe, CART algoritam je relativno robustan u slučaju manjka podataka (Breiman, 1984), omogućava jednostavnu interpretaciju i obradu dobijenih rezultata, statistički selektuje najznačajnija obeležja i omogućava kombinovanje kategorijskih (npr. identitet fonema) i numeričkih vrednosti (npr. trajanje fonema) obeležja.

Modelovanje trajanja govornih segmenata primenom CART tehnike podrazumeva upotrebu regresionog stabla za predviđanje trajanja datog govornog segmenta koji je u bazi predstavljen preko odgovarajućeg vektora obeležja. Formiranje pomenutog stabla sastoji se od nekoliko koraka: formiranje seta pitanja i izbor najboljeg pitanja na osnovu kojeg se vrši podela u datom čvoru; izbor kriterijuma za prestanak podele u nekom čvoru, odnosno proglašenje datog čvora za terminalni čvor (list) stabla; procena vrednosti u datom čvoru.

Neka je svaki od N podataka za obuku u bazi predstavljen preko odgovarajućeg vektora obeležja u formi:

$$X = (x_1, x_2, ..., x_M) \tag{6.1}$$

Za slučaj predviđanja trajanja x_1 može biti na primer način artikulacije prethodnog konsonanta, x_2 redni broj datog segmenta od kraja reči, itd. Treba zapaziti da elementi vektora obeležja mogu biti kategorijskog tipa, tj. mogu uzeti jednu od vrednosti iz konačnog neuređenog skupa (npr. način artikulacije konsonanta) ili numeričkog tipa, vrednost koju uzimaju je proizvoljan realan broj (npr. broj segmenata od kraja reči). U zavisnosti od tipa promenljive x_i formira se set pitanja Q:

1. Ako je nezavisna promenljiva x_i kategorijskog tipa, odnosno $x_i \in \{c_1, c_2, ..., c_K\} = C$ tada Q sadrži sva pitanja sledećeg oblika:
   $\{da\ li\ x_i \in A?\}$, $\forall A \subset C$

2. Ako je nezavisna promenljiva x_i numeričkog tipa, odnosno $-\infty < x_i < \infty$ tada Q sadrži sva pitanja sledećeg oblika:
   $\{da\ li\ x_i \le k?\}$, $\forall k$

Nakon formiranja celokupnog seta mogućih pitanja Q potrebno je pronaći najbolje pitanje za dati čvor, odnosno ono pitanje koje daje najbolju podelu podataka u datom čvoru.

Kod primene regresionih stabala, odnosno problema čije se rešavanje svodi na predviđanje kontinualne vrednosti najčešće primenjivani kriterijum podele jeste srednja kvadratna greška. Neka je Y stvarna vrednost trajanja nekog govornog segmenta u bazi predstavljenog preko vektora obeležja $\mathbf{X}$, tada se ukupna greška predikcije u čvoru t definiše kao:

$$E(t) = \sum_{\mathbf{X} \in t} |Y - d(\mathbf{X})|^2 \tag{6.2}$$

gde je $d(\mathbf{X})$ predviđena vrednost trajanja.

U skupu mogućih pitanja Q potrebno je pronaći pitanje koje najviše umanjuje kvadratnu grešku, odnosno pitanje q^* koje maksimizuje:

$$\Delta E_t(q) = E(t) - (E(l) + E(r)) \tag{6.3}$$

gde su l i r čvorovi koji se dobijaju nakon podele čvora t.

Za čvor t očekivana kvadratna greška definiše se kao:

$$V(t) = E\left\{ \sum_{\mathbf{X} \in t} |Y - d(\mathbf{X})|^2 \right\} = \frac{1}{N(t)} \sum_{\mathbf{X} \in t} |Y - d(\mathbf{X})|^2 \tag{6.4}$$

gde je $N(t)$ ukupan broj podataka koji se nalaze u datom čvoru t.

Može se zapaziti da $V(t)$ zapravo predstavlja varijansu procene trajanja ukoliko je $d(\mathbf{X})$ srednja vrednost trajanja svih govornih segmenata koji se nalaze u čvoru t.

Ponderisana kvadratna greška $\overline{V}(t)$ u čvoru t definiše se kao:

$$\overline{V}(t) = V(t) \cdot P(t) = (\frac{1}{N(t)} \sum_{\mathbf{X} \in t} \left| Y - d(\mathbf{X}) \right|^2) \cdot P(t) \tag{6.5}$$

gde je $P(t)$ odnos broja podataka u čvoru t i ukupnog broja podataka.

Konačno, kriterijum podele u nekom čvoru t može se napisati kao:

$$\Delta\overline{V}_t(q) = \overline{V}(t) - (\overline{V}(r) + \overline{V}(\mathrm{l})) \tag{6.6}$$

pri čemu je potrebno pronaći pitanje q koje minimizuje varijansu predikcije nakon podele čvora t na čvorove l i r.

Za dati skup pitanja Q i kriterijum podele $\overline{V}_t(q)$ proces formiranja stabla počinje od stabla koje sadrži samo koren, odnosno čvor u kome se nalazi svih N podataka za obuku iz baze. U svakom čvoru stabla za svaki element vektora obeležja x_i $i = 1,...,M$ algoritam pronalazi najbolje pitanje iz datog skupa Q koristeći odgovarajući kriterijum podele. Nakon toga, od ukupno M izabranih pitanja bira se najbolje među njima, odnosno vrši se izbor najznačajnijeg obeležja u datom čvoru. Opisani postupak se ponavlja za svaki novodobijeni čvor sve dok ne bude zadovoljen jedan od sledećih uslova:

1. nakon podele maksimalno smanjenje varijanse je ispod unapred utvrđenog praga β, t.j.:
$$\max_{q \in Q} \Delta\overline{V}_t(q) < \beta \tag{6.7}$$
2. broj podataka koji se nalaze u datom čvoru t je manji od unapred utvrđenog praga α.

Ukoliko je u nekom čvoru t zadovoljen jedan od prethodno navedenih uslova ne vrši se dalja podela u tom čvoru, odnosno čvor se proglašava za terminalni čvor stabla. Algoritam se završava kada je svaki čvor stabla proglašen za terminalni. Po završetku faze formiranja stabla zadovoljenjem nekih od prethodno navedenih uslova obično se dobija veliko stablo $T_{\max}$ koje može biti formirano striktno prema podacima koji su korišćeni u fazi obuke i takvo stablo nema sposobnost generalizacije, odnosno neće pokazati dobre performanse u slučaju primene nad podacima koji nisu korišćeni u fazi obuke. Stoga, potrebno je pronaći stablo optimalne veličine i izbeći *overfitting* podataka. U literaturi se navodi da je bilo niz pokušaja za prevazilaženje ovog problema među kojima se kao najbolje rešenje izdvaja Breimanov postupak koji se sastoji od nekoliko koraka: 1) formira se sekvenca podstabala $T_{\max} \supseteq ... \supseteq T_k \supseteq ... \supseteq T_K = t_1$ 2) za svako podstablo procenjuje se stopa greške 3) bira se stablo sa najmanjom stopom greške, odnosno

stablo optimalne veličine (Breiman, 1984). Opisani postupak naziva se potkresivanje stabla (engl. *cost-complexity pruning*). Prilikom formiranja sekvence podstabala koja se dobijaju odstranjivanjem pojedinih grana parametar kompleksnosti α varira od 0 (za $T_{\max}$) do ∞ (za podstablo koje sadrži samo koren) tako da je zadovoljen uslov:

$$\min_T \left[\sigma^2(T) + \alpha \cdot |T|\right] \tag{6.8}$$

gde je: $\sigma^2(T)$ varijansa greške predikcije za dato podstablo

$|T|$ broj terminalnih čvorova podstabla

U cilju procene stope greške podstabla se testiraju na podacima koji nisu korišćeni u fazi obuke. Procedura koja se najčešće primenjuje za procenu naziva se ukrštena validacija (engl. *cross-validation*). Naime, ukupna količina raspoloživih podataka podeli se na deset međusobno disjunktnih podskupova na kojima se vrši testiranje podstabala koja su formirana na osnovu preostalih 9/10 podataka. S obzirom na to da se postupak testiranja ponavlja deset puta za svako podstablo se izračunava prosečna varijansa. Ukoliko se varijansa dobijena na ovaj način posmatra kao funkcija veličine stabla, tada će za stablo određene veličine biti dostignut minimum varijanse i takvo stablo smatra se stablom optimalne veličine, jer dalje povećanje veličine stabla povećava varijansu.

Primenom CART tehnike formira se binarno stablo koje u svakom čvoru sadrži da/ne pitanje o nekom obeležju, odnosno faktoru koji utiče na trajanje govornog segmenta. Predviđanje trajanja govornog segmenta vrši se prolaskom kroz stablo odluke, od korena do lista stabla, prolazeći kroz unutrašnje čvorove stabla onom putanjom koja se formira u zavisnosti od zadovoljenja određenog uslova o vrednostima obeležja u svakom od unutrašnjih čvorova. List stabla sadrži predviđenu vrednost trajanja datog govornog segmenta.

Regresiono stablo predstavlja specijalan slučaj modelskog stabla. Jedina razlika između regresionog i modelskog stabla leži u činjenici da modelsko stablo u svakom čvoru sadrži linearan regresioni model, zasnovan na nekim vrednostima atributa, umesto konstantnu vrednost. Primenom linearnog regresionog modela u listu stabla dobija se predviđena vrednost trajanja datog govornog segmenta.

Pored regresionog i modelskog stabla u procesu modelovanja trajanja glasova korišćen je još jedan algoritam baziran na stablima odluke. To je REP (engl. *Reduced Error Pruning*) Trees algoritam koji je razvijen u cilju dobijanja optimalnog stabla, što podrazumeva pronalaženje stabla najmanje veličine uz ostvarivanje minimalne greške. Kod ovog algoritma takođe se koriste različiti skupovi podataka za formiranje i potkresivanje stabla (engl. *pruning)*, pri čemu je

međusoban odnos količine podataka u ova dva skupa jedan od parametara algoritma. Detaljniji opis REPTrees algoritma može se pronaći u (Kaariainen & Malinen, 2004).

6.2 LINEARNA REGRESIJA

Linearna regresija (Witten & Frank, 2005) predstavlja jednu od metoda koje se primenjuju u svrhu predviđanja određene numeričke vrednosti. Ova veoma jednostavna metoda ujedno je i najstarija metoda regresione analize, detaljno proučavana i dugi niz godina se intenzivno koristi u mnogim praktičnim primenama. Osnovna ideja ovog algoritma jeste da se zavisna promenljiva, odnosno vrednost koja se predviđa predstavi kao linearna kombinacija atributa koji se uzimaju u obzir prilikom formiranja prediktivnog modela i od kojih zavisi vrednost koja se predviđa, pri čemu je svaki od atributa ponderisan odgovarajućim težinskim faktorom. Stoga, traženu zavisnu promenljivu moguće je predstaviti kao:

$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k \tag{6.9}$$

gde je: x vrednost koja se predviđa, odnosno trajanje fonema u slučaju modelovanja trajanja

$a_1, a_2, \ldots, a_k$ faktori koji utiču na trajanje fonema

$w_0, w_1, \ldots, w_k$ odgovarajući težinski faktori

Težinski faktori se određuju na osnovu podataka za obuku koji se nalaze u govornoj bazi, a koji se potom koriste za predviđanje potpuno novih podataka.

Predviđenu vrednost trajanja prvog fonema u bazi moguće je predstaviti kao:

$$x = w_0 a_0^{(1)} + w_1 a_1^{(1)} + w_2 a_2^{(1)} + \ldots + w_k a_k^{(1)} = \sum_{j=0}^{k} w_j a_j^{(1)} \tag{6.10}$$

Prilikom određivanja koeficijenata w_j kojih ukupno ima $k+1$primenjuje se metod najmanjih kvadrata, odnosno potrebno je minimizovati sumu kvadrata razlike između stvarne i predviđene vrednosti trajanja za svaki fonem koji se nalazi u bazi.

Ukoliko u bazi postoji n fonema, pri čemu je *i*-ti fonem označen superskriptom (i) tada se suma kvadrata razlika može predstaviti u obliku:

$$\sum_{i=1}^{n} (x^{(i)} - \sum_{j=0}^{k} w_j a_j^{(i)})^2 \tag{6.11}$$

gde je razlika u zagradi razlika između stvarne i predviđene vrednosti trajanja *i*-tog fonema u bazi. Izborom odgovarajućih koeficijenata w_j suma kvadrata razlika se minimizuje.

6.3 META ALGORITMI

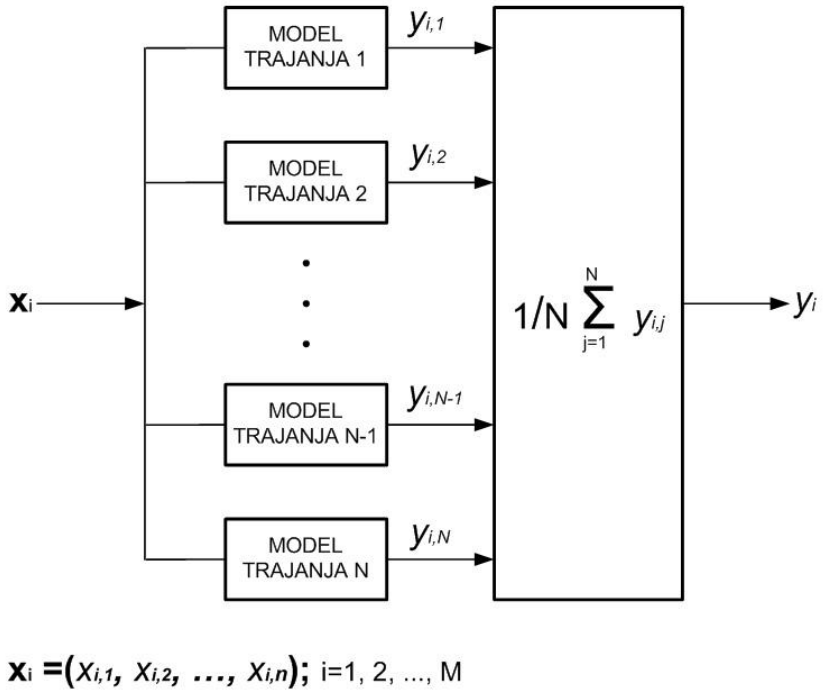
Meta algoritmi se baziraju na kombinovanju više različitih modela, odnosno primeni metoda automatskog učenja na meta podacima (Witten & Frank, 2005), koji predstavljaju podatke dobijene na osnovu nekih drugih podataka. Ovi algoritmi primenjuju se prilikom rešavanja problema numeričke predikcije i veoma često doprinose poboljšanju prediktivnih performansi u odnosu na individualne modele. Generalni nedostatak meta algoritama jeste otežana analiza jer mogu sadržati desetine ili čak i stotine individualnih modela što onemogućava jednostavno utvrđivanje doprinosa pojedinih faktora u poboljšanju prediktivnih performansi modela. Među najznačajnije meta algoritme ubrajaju se aditivna regresija, *bagging* i *stacking* koji će u nastavku biti ukratko opisani.

6.3.1 Aditivna regresija

Aditivna regresija (Stone, 1985) predstavlja jedan od meta algoritama čijom primenom se poboljšavaju performanse standardne regresione tehnike. Ovo poboljšanje ostvaruje se generisanjem predikcija sumirajući doprinose više individualnih modela, pri čemu se teži formiranju ansambla individualnih modela, koji optimizuje prediktivne performanse u skladu sa odgovarajućim kriterijumom. U fazi obuke, u svakoj iteraciji, formira se standardno regresiono stablo koristeći reziduale prethodnog stabla kao podatke za obuku. U slučaju modelovanja trajanja reziduali predstavljaju grešku predikcije, odnosno razliku između stvarne i predviđene vrednosti trajanja fonema. U svakoj sledećoj iteraciji vrši se predviđanje reziduala iz prethodne iteracije i na taj način se automatski smanjuje greška iz prethodne iteracije. Parametri algoritma koje je moguće specificirati su maksimalan broj iteracija i parametar koji određuje brzinu obučavanja v (engl. *shrinkage parameter*). Parametar v je unapred specificirana konstantna vrednost između 0 i 1 kojom se množi predviđena vrednost čime se smanjuje mogućnost da nastane *overfitting* podataka. Na ovaj način se naravno povećava i broj iteracija potrebnih za dobijanje optimalnog modela. U istraživanjima u okviru ove disertacije kao individualni modeli korišćeni su M5PR i REPTrees (Hall et al., 2009).

6.3.2 *Bagging*

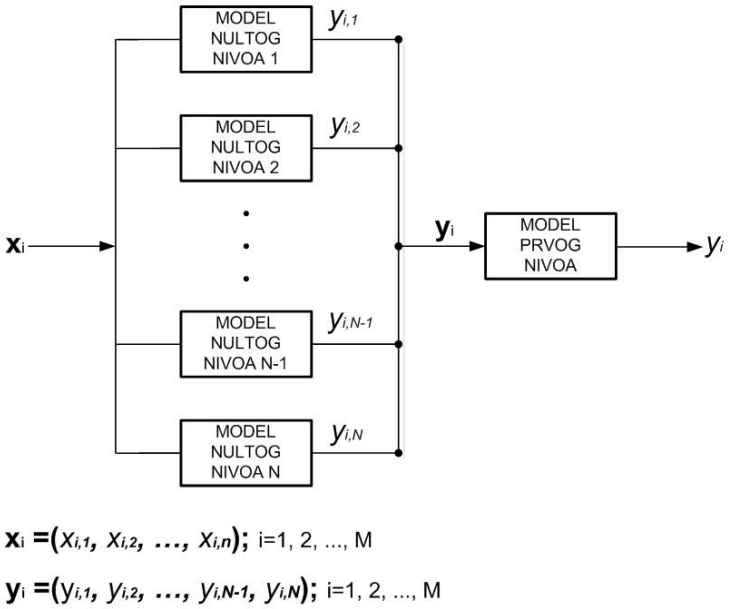
*Bagging (*engl. *Bootstrap aggregating*) algoritam (Breiman, 1996) je takođe jedan od meta algoritama koji se može primeniti prilikom modelovanja trajanja fonema. Ovaj algoritam baziran je na kombinovanju predikcija koje daju različiti individualni modeli, a konačna predviđena vrednost predstavlja srednju vrednost pojedinačnih predikcija svakog modela. Opšta struktura algoritma prikazana je na slici 6.1. Ukupan broj individualnih modela jeste parametar koji se specificira unapred. Svaki model se razvija koristeći različit skup podataka za obuku koji je iste veličine kao i polazni originalni skup podataka za obuku a dobijen je odabirom podataka iz originalnog skupa na slučajan način. Stoga, može se desiti da neki podatak bude ponovljen više puta u novodobijenom skupu podataka dok neki drugi može biti izostavljen. Teoretski se može pokazati da usrednjavanjem predikcija individualnih modela, koji su razvijeni primenom nezavisnih različitih skupova podataka za obuku, očekivana vrednost srednje kvadratne greške se uvek smanjuje (Witten & Frank, 2005). U istraživanjima u okviru ove disertacije kao individualni modeli korišćeni su M5PR i REPTrees algoritmi (Hall et al., 2009)



Slika 6.1 – Bagging algoritam

6.3.3 Stacking

Stacking algoritam (Witten & Frank, 2005) predstavlja još jedan od mogućih načina za kombinovanje individualnih modela. Ovaj metod, iako nastao pre *bagging* metoda, mnogo ređe se primenjuje delimično zbog otežane teoretske analize a delom zbog činjenice da je osnovnu ideju moguće realizovati na mnogo različitih načina. Opšta struktura *stacking* algoritma prikazana je na slici 6.2. Za razliku od *bagging* metode, gde se kombinuje više individualnih modela istog tipa, kod *stacking* metode postoji mogućnost kombinovanja više modela različitog tipa. Najbolji način za kombinovanje rezultata koje daju individualni modeli utvrđuje se primenom posebne metode automatskog učenja koja zapravo predstavlja meta obučavač (engl. *metalearner*), odnosno meta model. Meta model, odnosno model prvog nivoa imaće onoliko atributa koliko ima individualnih modela nultog nivoa, dok su vrednosti atributa na prvom nivou upravo predikcije koje daju individualni modeli nultog nivoa. Prilikom obučavanja modela nultog nivoa obično se primenjuje postupak ukrštene validacije, čime se obezbeđuje korišćenje svih podataka za obuku nultog nivoa i prilikom obučavanja modela prvog nivoa. Mnogobrojne metode automatskog učenja mogu biti primenjene za razvoj modela prvog nivoa. S obzirom na to da se najveći deo obučavanja podataka obavlja na nultom nivou poželjno je da se na prvom nivou primenjuje neki jednostavan algoritam kao na primer linearna regresija ili modelsko stablo.



Slika 6.2 – Stacking algoritam

7. RAZVOJ MODELA TRAJANJA GLASOVA PRIMENOM SOFTVERSKOG PAKETA WEKA

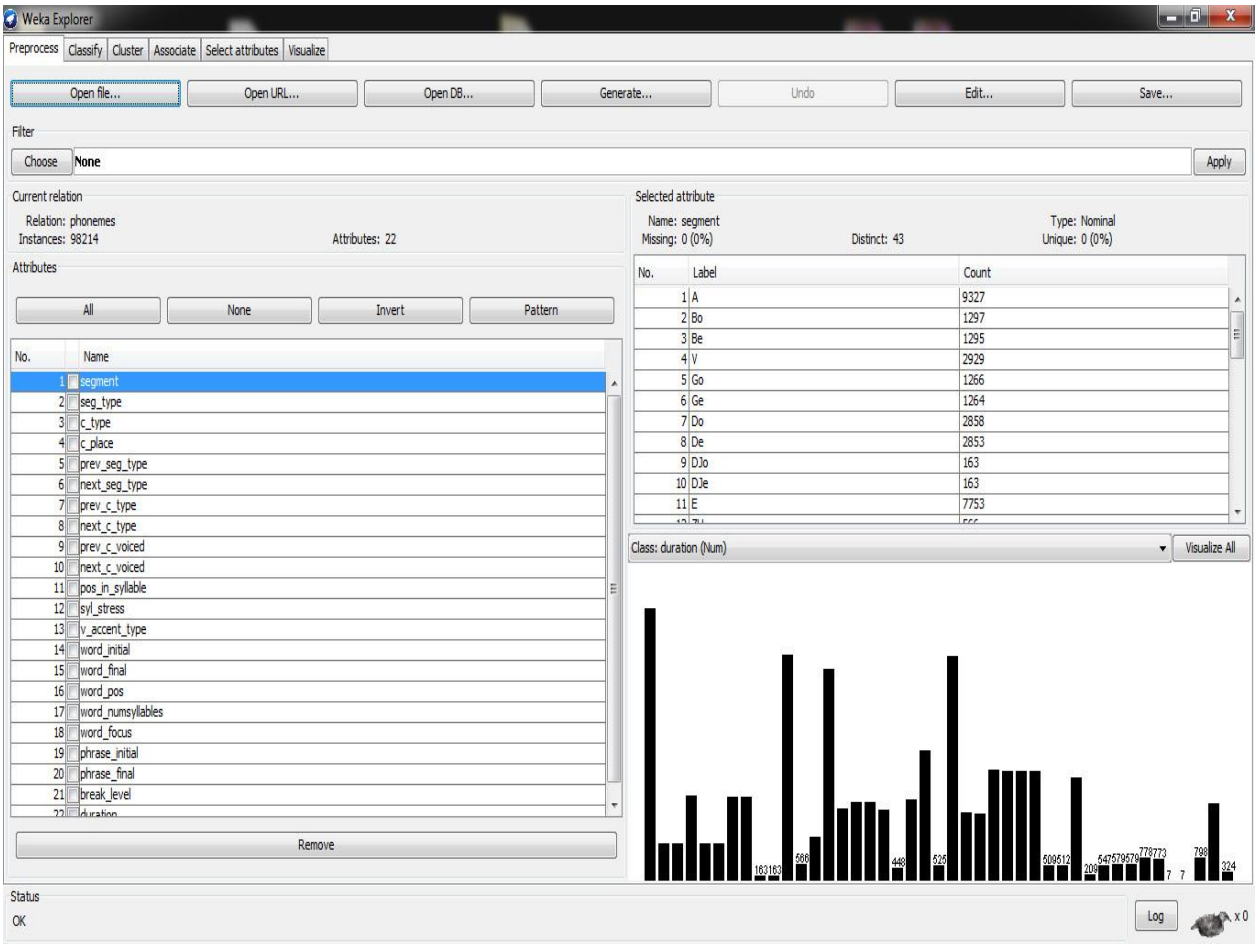
7.1 OPIS SOFTVERSKOG PAKETA WEKA

Prilikom modelovanja trajanja glasova u srpskom jeziku u okviru ove disertacije korišćen je softverski paket WEKA (engl. *Waikato Environment for Knowledge Analysis*) (Hall et al., 2009). WEKA je softver sa otvorenim kodom razvijen u programskog jeziku Java na Univerzitetu Waikato na Novom Zelandu, koji sadrži mnogobrojne algoritme automatskog učenja a namenjen je rešavanju različitih *data mining* problema. Algoritmi mogu biti direktno primenjeni na odgovarajući set podataka ili pozvani iz određenog Java programskog koda. WEKA sadrži alate koji omogućavaju klasifikaciju, regresiju ili klasterovanje podataka, utvrđivanje odgovarajućih pravila, pretprocesiranje ulaznih podataka, statističku evaluaciju modela dobijenog primenom metode automatskog učenja, kao i vizuelizaciju ulaznih podataka i rezultata obučavanja. Primena metoda automatskog učenja omogućava automatsku analizu velike količine podataka i utvrđivanje najrelevantnijih informacija koje mogu biti upotrebljene u različitim prediktivnim procesima. Primena ovih metoda takođe može doprineti bržem i tačnijem donošenju odluka. S druge strane, simbolika naziva leži u činjenici da je Weka ptica radoznale prirode bez mogućnosti letenja, koja predstavlja autohtonu vrstu na ostrvima Novog Zelanda i može se pronaći samo tamo.

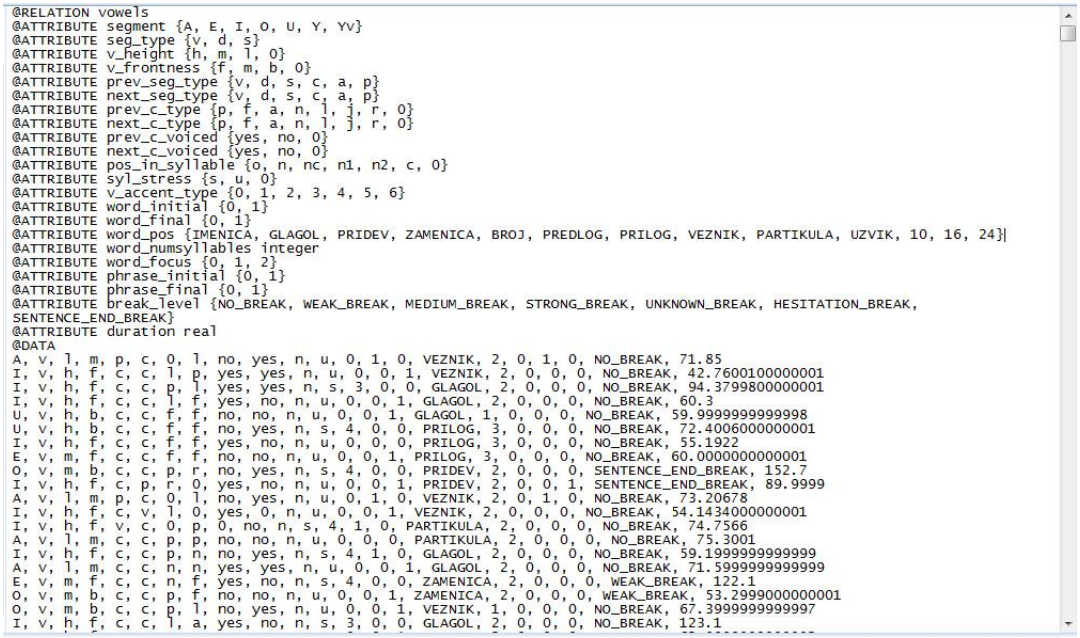
Slika 7.1 – WEKA korisnički interfejs

Softver WEKA distribuira se u skladu sa *GNU General Public Licence*, može da radi na gotovo svakoj platformi a testiran je u različitim operativnim sistemima kao što su Linux, Windows, Macintosh, pa čak i PDA (engl. *Personal Digital Assistant*). WEKA poseduje četiri različita korisnička interfejsa, pri čemu su Explorer, Knowledge Flow i Experimenter grafički korisnički interfejsi dok upotreba četvrtog interfejsa podrazumeva unošenje tekstualnih komandi sa komandne linije (slika 7.1).

Prilikom razvoja modela trajanja glasova u srpskom jeziku korišćen je grafički korisnički interfejs Explorer koji ujedno predstavlja najčešće primenjivani interfejs softverskog paketa WEKA. Na slici 7.2 prikazan je WEKA Explorer interfejs i učitavanje podataka iz odgovarajućeg fajla. Ulazni podaci moraju biti u ARFF (engl. *Attribute-Relation File Format*) formatu. ARFF fajl je ASCII tekstualni fajl u kome su na odgovarajući način smešteni podaci koji su opisani određenim skupom atributa. Svaki ARFF fajl sadrži dva odeljka, pri čemu prvi predstavlja zaglavlje (engl. *header section*) a u drugom su smešteni podaci (engl. *data section*). Na slici 7.3 dat je primer ARFF fajla.



Slika 7.2 – WEKA Explorer interfejs



```
@RELATION vowels
@ATTRIBUTE segment {A, E, I, O, U, Y, Yv}
@ATTRIBUTE seg_type {v, d, s}
@ATTRIBUTE v_height {h, m, l, 0}
@ATTRIBUTE v_frontness {f, m, b, 0}
@ATTRIBUTE prev_seg_type {v, d, s, c, a, p}
@ATTRIBUTE next_seg_type {v, d, s, c, a, p}
@ATTRIBUTE prev_c_type {p, f, a, n, l, j, r, 0}
@ATTRIBUTE next_c_type {p, f, a, n, l, j, r, 0}
@ATTRIBUTE prev_c_voiced {yes, no, 0}
@ATTRIBUTE next_c_voiced {yes, no, 0}
@ATTRIBUTE pos_in_syllable {o, n, nc, n1, n2, c, 0}
@ATTRIBUTE syl_stress {s, u, 0}
@ATTRIBUTE v_accent_type {0, 1, 2, 3, 4, 5, 6}
@ATTRIBUTE word_initial {0, 1}
@ATTRIBUTE word_final {0, 1}
@ATTRIBUTE word_pos {IMENICA, GLAGOL, PRIDEV, ZAMENICA, BROJ, PREDLOG, PRILOG, VEZNIK, PARTIKULA, UZVIK, 10, 16, 24}|
@ATTRIBUTE word_numsyllables integer
@ATTRIBUTE word_focus {0, 1, 2}
@ATTRIBUTE phrase_initial {0, 1}
@ATTRIBUTE phrase_final {0, 1}
@ATTRIBUTE break_level {NO_BREAK, WEAK_BREAK, MEDIUM_BREAK, STRONG_BREAK, UNKNOWN_BREAK, HESITATION_BREAK,
SENTENCE_END_BREAK}
@ATTRIBUTE duration real
@DATA
A, v, l, m, p, c, 0, l, no, yes, n, u, 0, 1, 0, VEZNIK, 2, 0, 1, 0, NO_BREAK, 71.85
I, v, h, f, c, c, l, p, yes, yes, n, u, 0, 0, 1, VEZNIK, 2, 0, 0, 0, NO_BREAK, 42.7600100000001
I, v, h, f, c, c, p, l, yes, yes, n, s, 3, 0, 0, GLAGOL, 2, 0, 0, 0, NO_BREAK, 94.3799800000001
I, v, h, f, c, c, l, f, yes, no, n, u, 0, 0, 1, GLAGOL, 2, 0, 0, 0, NO_BREAK, 60.3
U, v, h, b, c, c, f, f, no, no, n, u, 0, 0, 1, GLAGOL, 1, 0, 0, 0, NO_BREAK, 59.9999999999998
U, v, h, b, c, c, f, f, no, yes, n, s, 4, 0, 0, PRILOG, 3, 0, 0, 0, NO_BREAK, 72.4006000000001
I, v, h, f, c, c, f, f, yes, no, n, u, 0, 0, 0, PRILOG, 3, 0, 0, 0, NO_BREAK, 55.1922
E, v, m, f, c, c, f, f, no, no, n, u, 0, 0, 1, PRILOG, 3, 0, 0, 0, NO_BREAK, 60.0000000000001
O, v, m, b, c, c, p, r, no, yes, n, s, 4, 0, 0, PRIDEV, 2, 0, 0, 0, SENTENCE_END_BREAK, 152.7
I, v, h, f, c, p, r, 0, yes, no, n, u, 0, 0, 1, PRIDEV, 2, 0, 0, 1, SENTENCE_END_BREAK, 89.9999
A, v, l, m, p, c, 0, l, no, yes, n, u, 0, 1, 0, VEZNIK, 2, 0, 1, 0, NO_BREAK, 73.20678
I, v, h, f, c, v, l, 0, yes, 0, n, u, 0, 0, 1, VEZNIK, 2, 0, 0, 0, NO_BREAK, 54.1434000000001
I, v, h, f, v, c, 0, p, 0, no, n, s, 4, 1, 0, PARTIKULA, 2, 0, 0, 0, NO_BREAK, 74.7566
A, v, l, m, c, c, p, p, no, no, n, u, 0, 0, 0, PARTIKULA, 2, 0, 0, 0, NO_BREAK, 75.3001
I, v, h, f, c, c, p, n, no, yes, n, s, 4, 1, 0, GLAGOL, 2, 0, 0, 0, NO_BREAK, 59.1999999999999
A, v, l, m, c, c, n, n, yes, yes, n, u, 0, 0, 1, GLAGOL, 2, 0, 0, 0, NO_BREAK, 71.5999999999999
E, v, m, f, c, c, n, f, yes, no, n, s, 4, 0, 0, ZAMENICA, 2, 0, 0, 0, WEAK_BREAK, 122.1
O, v, m, b, c, c, p, f, no, no, n, u, 0, 0, 1, ZAMENICA, 2, 0, 0, 0, WEAK_BREAK, 53.2999000000001
O, v, m, b, c, c, p, l, no, yes, n, u, 0, 0, 1, VEZNIK, 1, 0, 0, 0, NO_BREAK, 67.3999999999997
I, v, h, f, c, c, l, a, yes, no, n, s, 3, 0, 0, GLAGOL, 2, 0, 0, 0, NO_BREAK, 123.1
```

Slika 7.3 – Primer ARFF fajla

Zaglavlje ARFF fajla sadrži naziv odgovarajuće relacije, listu atributa, kao i njihove tipove i moguće vrednosti. U drugom odeljku ARFF fajla u kolonama se nalaze vrednosti odgovarajućih atributa za svaki podatak. Ključne reči @RELATION, @ATTRIBUTE i @DATA su deklaracione i ne postoji razlika u upotrebljenoj veličini slova. U okviru zaglavlja deklarišu se određena relacija i odgovarajući atributi. Deklaracija relacije ima sledeći format:

@RELATION <relation-name>

pri čemu je <relation-name> string. Ukoliko naziv relacije sadrži prazna polja mora biti naveden pod jednostrukim znacima navoda. Nakon deklarisanja relacije sledi deo u kom su deklarisani atributi, pri čemu redosled određenog atributa ukazuje na položaj kolone u kojoj se nalaze vrednosti datog atributa u okviru odeljka u kom su smešteni podaci. Format deklaracije atributa je sledeći:

@ATTRIBUTE <attribute-name> <datatype>

pri čemu <attribute-name> mora početi slovnim karakterom. Ukoliko naziv atributa sadrži prazna polja neophodno je da se celokupni naziv nađe pod jednostrukim znacima navoda.

WEKA podržava četiri tipa podataka, te stoga <datatype> može biti bilo koji od sledećih tipova:

- numeric
- <nominal-specification>
- string
- date [<data-format>]

Tipovi podataka *integer* i *real* se tretiraju kao *numeric*, stoga numerički atributi mogu imati realnu ili celobrojnu vrednost. Nominalne vrednosti atributa definišu se kao <nominal-specification>, odnosno sadrže listu mogućih vrednosti atributa:

{<nominal-name1>, <nominal-name2>, <nominal-name3>, …}

Na primer, skup mogućih vrednosti atributa segment definiše se na sledeći način:

@ATTRIBUTE segment {A, E, I, O, U, Y, Yv}

Ukoliko vrednosti atributa sadrže prazno polje moraju biti navedene pod jednostrukim znacima navoda.

Atributi tipa string deklarišu se kao:

@ATTRIBUTE <attribute-name> string

Ovaj tip podataka omogućava kreiranje atributa koji sadrže proizvoljne tekstualne vrednosti.

Atributi tipa date deklarišu se kao:

@ATTRIBUTE <name> date [<data-format>]

pri čemu je <name> naziv atributa, a [<data-format>] string koji specificira format u kom se datum prikazuje.

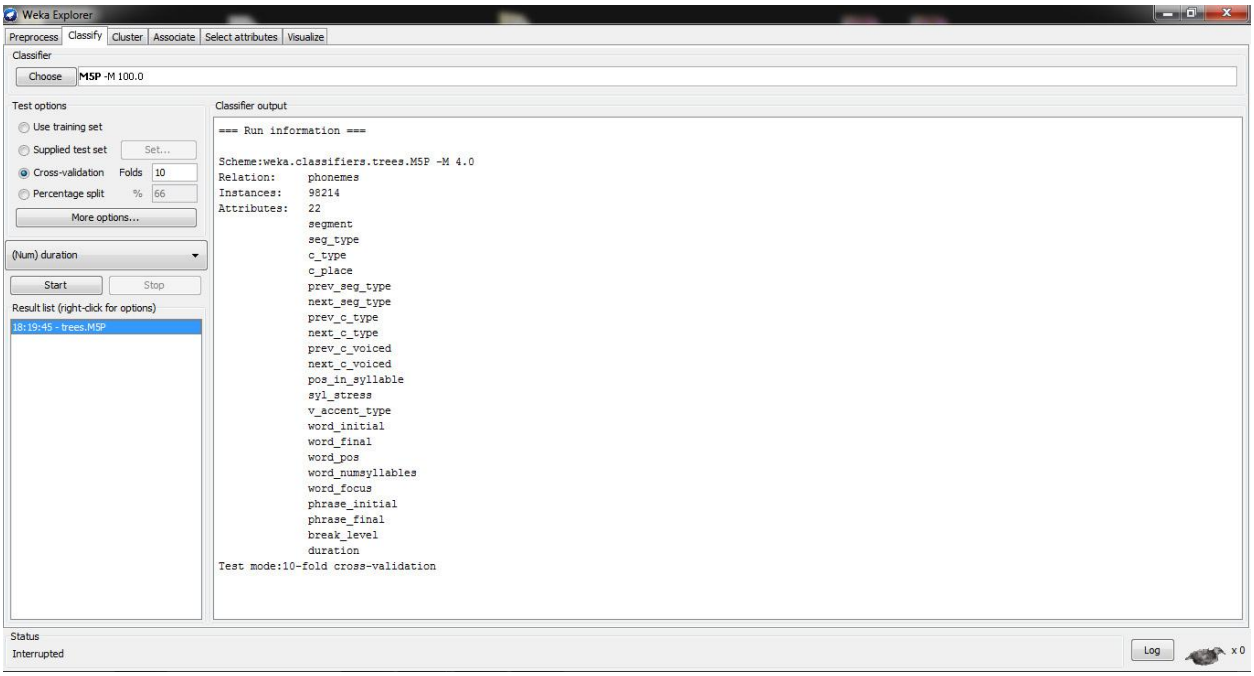
Prva linija u odeljku gde se nalaze podaci je deklaraciona a zatim slede linije u kojima su prikazani podaci. Format deklaracione linije je:

@DATA

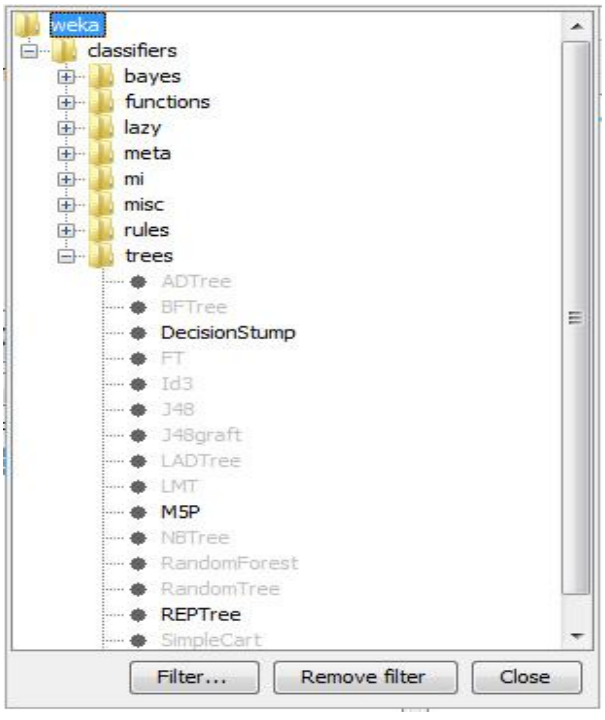
Svaki podatak predstavljen je u jednoj liniji pri čemu se vrednost određenog atributa nalazi u odgovarajućoj koloni. Vrednost atributa koji je deklarisan kao *n-ti* nalazi se u *n-toj* koloni.

Pored učitavanja ARFF fajlova WEKA omogućava i učitavanje fajlova u CSV (engl. *Comma Separated Values*) formatu i automatsku konverziju u ARFF format.

Nakon učitavanja podataka u ARFF formatu moguć je razvoj odgovarajućeg modela, pri čemu se klikom na *Classify* otvara prozor kao na slici 7.4.



Slika 7.4 – Izbor odgovarajućeg algoritma

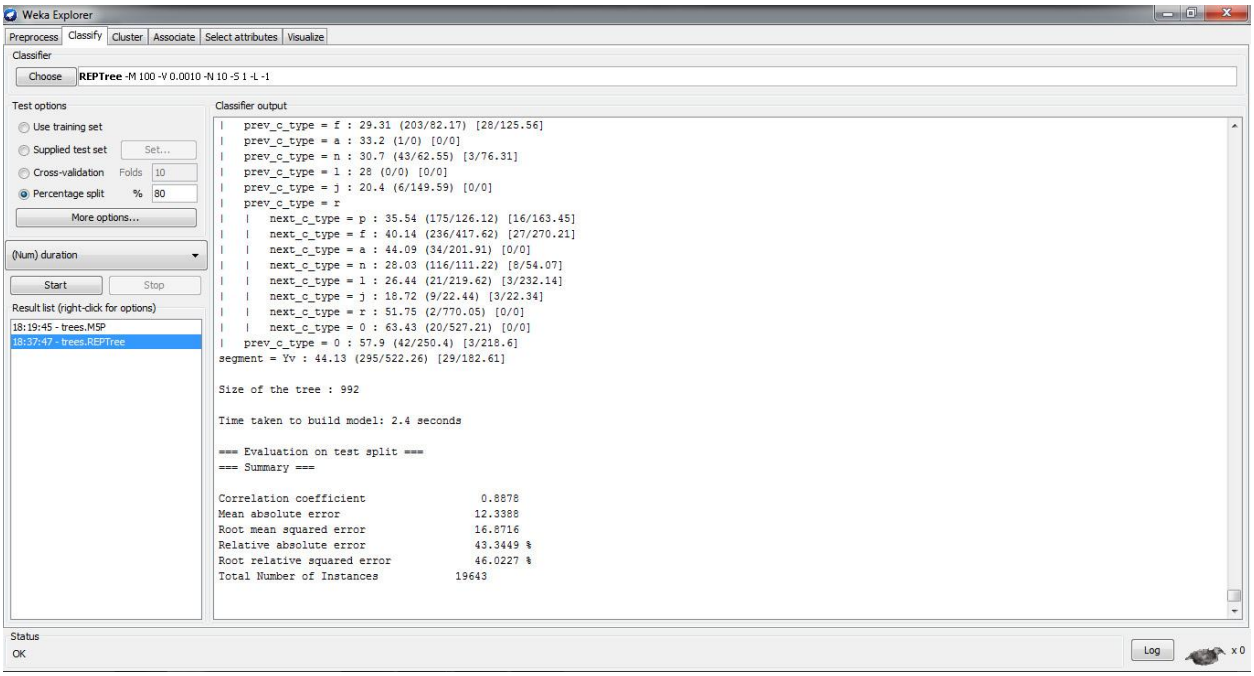


Slika 7.5 – WEKA algoritmi

Izbor odgovarajućeg algoritma vrši se klikom na *Choose* u odeljku *Classifier*. Na slici 7.5 prikazane su grupe algoritama koje su na raspolaganju u okviru softverskog paketa WEKA. Prilikom razvoja modela trajanja glasova u srpskom jeziku korišćeno je nekoliko algoritama kao što su M5P, M5PR i REPTree, koji pripadaju grupi *trees*, LinearRegression koji se nalazi u grupi *function*, kao i meta algoritmi AditiveRegression, Bagging i Stacking. Parametri koji su korišćeni prilikom razvoja određenog modela, kao i dobijeni rezultati biće prikazani u okviru ovog poglavlja.

WEKA nudi mogućnost ocene dobijenog modela na četiri različita načina. Objektivna evaluacija modela moguća je testiranjem modela na podacima koji su korišćeni prilikom razvoja modela, na određenom skupu podataka za testiranje, primenom postupka ukrštene validacije pri

čemu se ukupna količina podataka deli na određeni broj međusobno disjunktnih podskupova, kao i podelom skupa podataka na dva dela pri čemu se u svakom podskupu nalazi određeni procenat ukupne količine podataka. Na slici 7.6 su prikazani rezultati dobijeni nakon primene REPTree algoritma. Evaluacija dobijenog modela i poređenje sa drugim modelima biće u okviru ove disertacije vršeni na osnovu kvantitativnih pokazatelja kao što su RMSE (engl. *Root-Mean-Square Error*), MAE (engl. *Mean Absolute Error*) i CC (engl. *Correlation Coefficient*).



Slika 7.6 – Rezultati evaluacije REPTree modela

RMSE predstavlja veoma često korišćenu kvantitativnu meru za ocenu performansi razvijenog modela trajanja glasova, koja se definiše kao:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{7.1}$$

gde su: y_i stvarna vrednost trajanja *i-tog* fonema u ms

$\hat{y}_i$ predviđena vrednost trajanja *i-tog* fonema u ms

n ukupan broj fonema

MAE predstavlja srednju apsolutnu grešku između stvarne i predviđene vrednosti trajanja fonema i definiše se kao:

$$MAE = \frac{\sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|}{n} \tag{7.2}$$

gde su: y_i stvarna vrednost trajanja *i-tog* fonema u ms

$\hat{y}_i$ predviđena vrednost trajanja *i-tog* fonema u ms

n ukupan broj fonema

Koeficijent korelacije pokazuje statističku korelaciju između stvarnih i predviđenih vrednosti trajanja fonema i definiše se kao:

$$CC = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y} \tag{7.3}$$

gde su: X slučajna promenljiva koja predstavlja predviđene vrednosti trajanja fonema u ms

Y slučajna promenljiva koja predstavlja stvarne vrednosti trajanja fonema u ms

μ_X, μ_Y srednje vrednosti slučajnih promenljivih

σ_X, σ_Y standardne devijacije slučajnih promenljivih

7.2 EVALUACIJA I POREĐENJE MODELA TRAJANJA

Prilikom razvoja modela trajanja glasova u srpskom jeziku korišćeni su različiti algoritmi softverskog paketa WEKA (Hall et al., 2009) među koje se ubrajaju LinearRegression (LR), M5P, M5PR, REPTree, kao i meta algoritmi AditiveRegression (AR), Bagging (BAG) i Stacking (STACK). Pomenuti modeli razvijani su primenom odgovarajućeg algoritma na obimnom govornom korpusu koji sadrži 98214 glasova, odnosno 38543 vokala i 59671 konsonant. Broj pojavljivanja određenog fonema u govornoj bazi prikazan je na slici 5.1 u jednom od prethodnih poglavlja. Modeli trajanja razvijeni su za celokupan skup fonema u srpskom jeziku. Takođe, razvijeni su i posebni modeli za vokale, odnosno konsonante.

Evaluacija razvijenih modela trajanja glasova realizovana je pomoću postupka ukrštene validacije, pri čemu je ukupna količina podataka podeljena na deset međusobno disjunktnih podskupova što je veoma čest slučaj prilikom testiranja primenom ovog postupka. Poređenje razvijenih modela izvršeno je na osnovu kvantitativnih pokazatelja kao što su RMSE, MAE i CC između stvarne i predviđene vrednosti trajanja glasova. Ocena prediktivnih performansi modela urađena je takođe i na podacima koji nisu bili korišćeni u fazi obuke modela. U ovom slučaju celokupna govorna baza bila je podeljena na dva dela. Skup koji je korišćen za obuku modela sadrži 80% govorne baze, dok je testiranje modela vršeno na preostalih 20%.

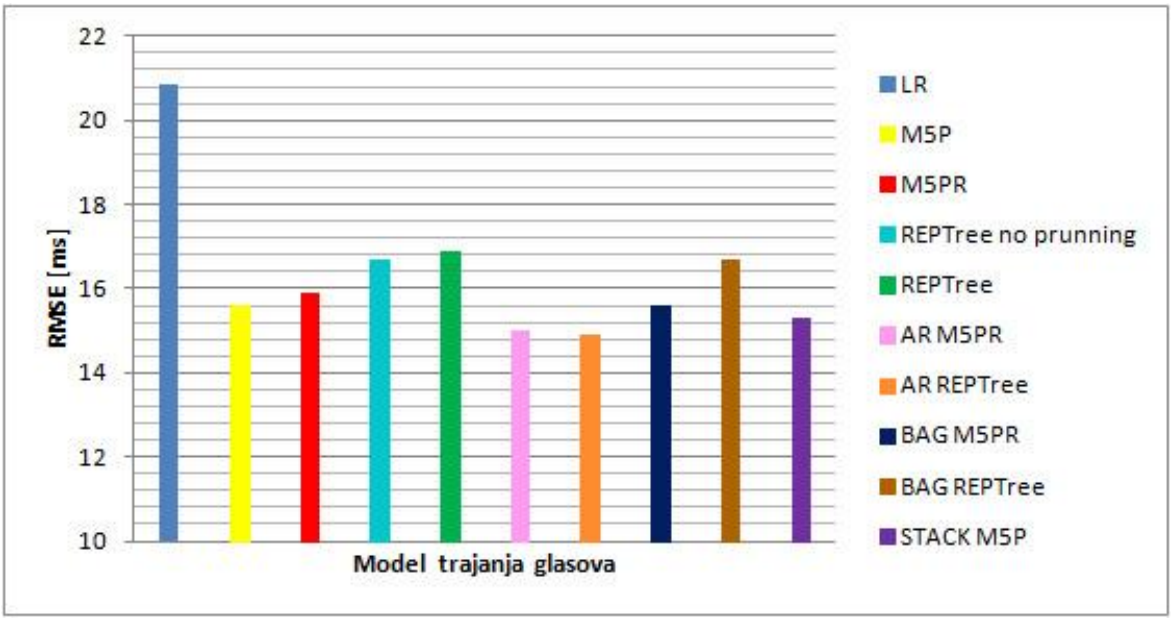
U tabeli 7.1 prikazani su rezultati dobijeni primenom različitih algoritama kao što su LinearRegression (LR), M5P, M5PR, REPTree, AditiveRegression (AR), Bagging (BAG) I Stacking (STACK) na celokupan skup fonema. REPTree algoritam primenjen je u dve varijante, odnosno sa i bez potkresivanja (engl. *pruning*) stabla. M5P i REPTree algoritmi korišćeni su kao bazični algoritmi kod AR i BAG meta algoritama. Kod aditivne regresije kada je osnovni algoritam bio M5P broj iteracija je deset a petnaest kada je osnovni algoritam bio REPTree, dok je u oba slučaja parameter v koji određuje brzinu obučavanja iznosio 0,5.

Model trajanja glasova (način testiranja)	**RMSE [ms]**	**MAE [ms]**	**CC**
LR (cross-validation)	20,8834	15,7481	0,8203
LR (20% test set)	20,8554	15,7226	0,8224
M5P (cross-validation)	15,5579	11,4724	0,9047
M5P (20% test set)	15,5996	11,4915	0,9050
M5PR (cross-validation)	15,8307	11,6490	0,9012
M5PR (20% test set)	15,9160	11,7058	0,9008
REPTree noprunning (cross-validation)	16,5317	12,0729	0,8917
REPTree noprunning (20% test set)	16,7069	12,1194	0,8901
REPTree (cross-validation)	16,7414	12,2118	0,8887
REPTree (20% test set)	16,8716	12,3388	0,8878
AR M5PR (cross-validation)	14,9304	11,0889	0,9126
AR M5PR (20% test set)	15,0027	11,1303	0,9125
AR REPTree (cross-validation)	14,8884	11,0655	0,9131
AR REPTree (20% test set)	14,9191	11,0746	0,9135
BAG M5PR (cross-validation)	15,5257	11,4526	0,9052
BAG M5PR (20% test set)	15,6014	11,4905	0,9050
BAG REPTree (cross-validation)	16,5482	12,0850	0,8914
BAG REPTree (20% test set)	16,7044	12,2007	0,8902
STACK M5P (cross-validation)	15,1423	11,2010	0,9100
STACK M5P (20% test set)	15,3170	11,3226	0,9085

Tabela7.1 – Prediktivne performanse modela trajanja glasova (dva načina testiranja)

Ukupan broj individualnih modela kod Bagging algoritma u oba slučaja bio je deset. LR, M5P, M5PR, REPTree bez potkresivanja i REPTree sa potkresivanjem stabla korišćeni su kao modeli nultog nivoa kod Stacking algoritma dok je meta obučavač bio M5P algoritam. Svi prethodno pomenuti modeli testirani su na dva različita načina. U slučaju kada su modeli testirani na podacima koji čine 20% ukupne govorne baze, a koji nisu bili korišćeni u fazi obuke, performanse modela su približno iste kao i kada je testiranje vršeno primenom postuka ukrštene validacije. Pomenuta činjenica važi za sve modele a njena važnost je u tome što ukazuje na dobre realne prediktivne performanse modela. Na osnovu rezultata prikazanih u tabeli 7.1 može se zapaziti da svi modeli poseduju zadovoljavajuće prediktivne performanse ostvarujući RMSE od 14,9191 do 20,8554 ms, MAE od 11,0746 do 15,7226 ms i CC od 0,8224 do 0,9135 u slučaju kada je testiranje modela vršeno na podacima koji nisu bili korišćeni u fazi obuke modela.

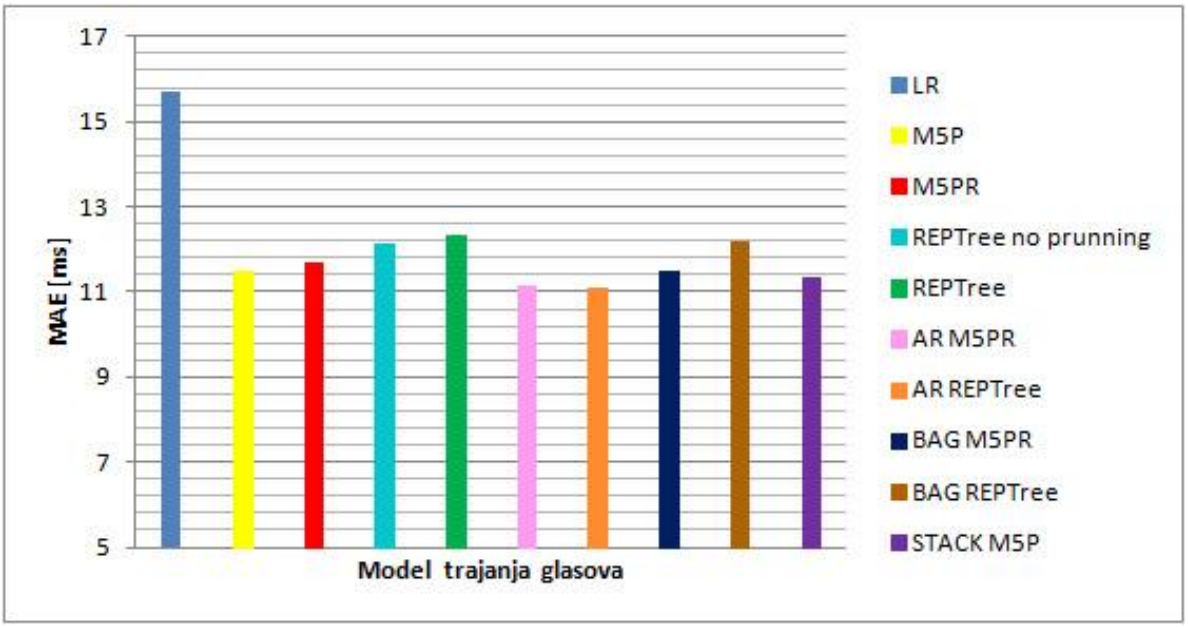
Na slikama 7.7, 7.8 i 7.9 histogramom su prikazani RMSE, MAE i CC svih razvijenih modela za celokupan skup fonema u srpskom jeziku. Na pomenutim slikama može se uočiti da najbolje performanse ostvaruje AR REPTree model. REPTree algoritam kao osnovni algoritam kod aditivne regresije daje bolje rezultate nego M5PR algoritam, dok se pokazalo da je M5PR algoritam bolji kao osnovni nego REPTree kod Bagging algoritma. Takođe, na histogramima se može primetiti da modeli dobijeni primenom nekog od meta algoritama poseduju bolje performanse nego LR model ili modeli zasnovani na primeni stabala odluke.



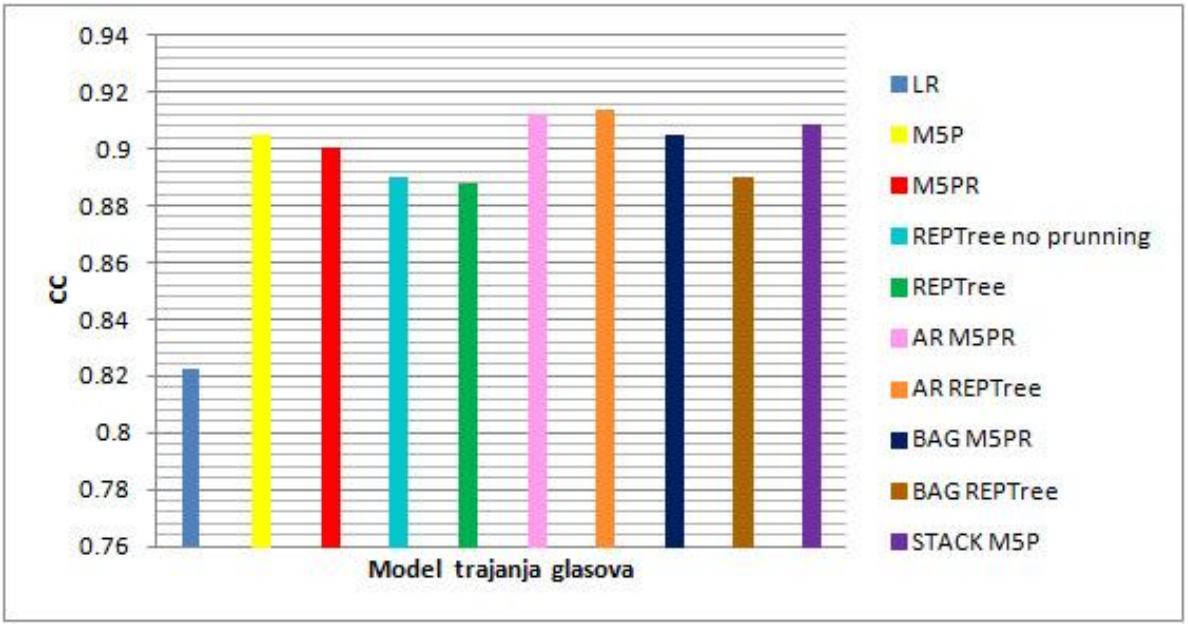
Slika 7.7 – RMSE kod različitih modela trajanja glasova

Performanse M5PR modela su neznatno lošije od prediktivnih performansi M5P modela što je od izuzetne važnosti, s obzirom da primena M5PR modela za predikciju trajanja glasova

smanjuje vreme predikcije iako je broj terminalnih čvorova veći nego kod M5P modela imajući u vidu da listovi stabla sadrže konstantnu vrednost koja ujedno predstavlja predviđenu vrednost trajanja datog glasa. Najlošije performanse u pogledu sva tri kvantitativna pokazatelja zapažaju se kod LR modela.



Slika 7.8 – MAE kod različitih modela trajanja glasova



Slika 7.9 – Koeficijent korelacije kod različitih modela trajanja glasova

U tabelama 7.2 i 7.3 date su vrednosti kvantitativnih pokazatelja kao što su RMSE, MAE i CC koji ukazuju na prediktivne performanse razvijenih modela trajanja konsonanata, odnosno

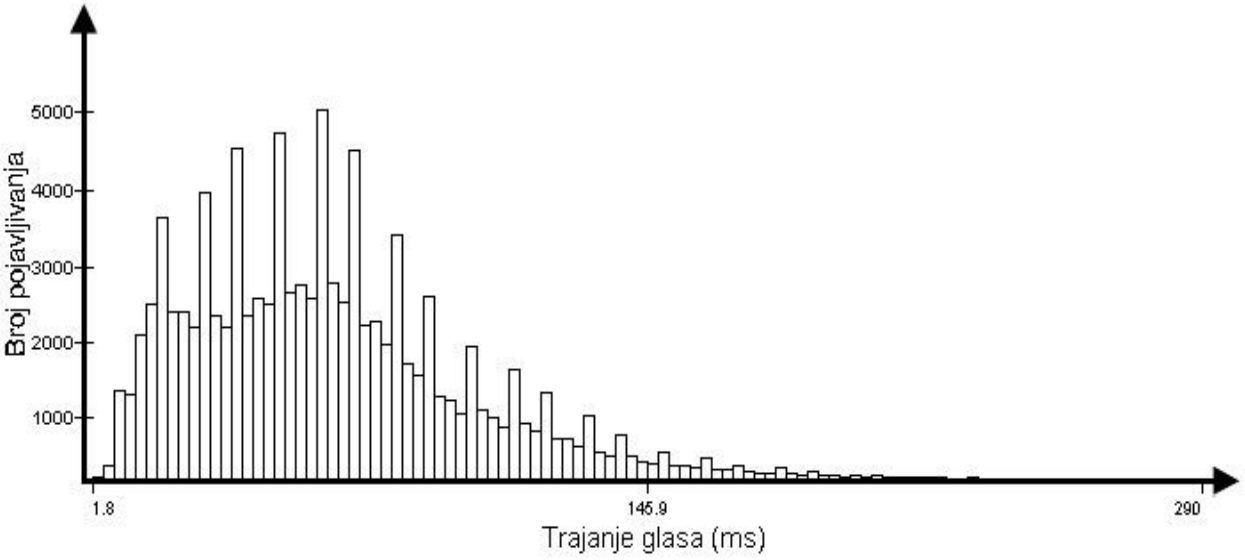
vokala. Prilikom razvoja modela trajanja konsonanata, kao i modela trajanja vokala korišćeni su već prethodno pomenuti algoritmi i odgovarajući skup za obuku, dok je testiranje modela izvršeno na odgovarajućem test skupu. Ukupan broj konsonanata u skupu za obuku, odnosno testiranje je 47737 i 11934, respektivno. Skupovi za obuku i testiranje modela trajanja vokala sadrže 30835 i 7708 vokala, respektivno. Rezultati prikazani u tabelama 7.2 i 7.3 ukazuju takođe na zadovoljavajuće prediktivne performanse razvijenih modela trajanja konsonanata, odnosno vokala. Modeli trajanja konsonanata ostvaruju RMSE od 13,8348 do 16,6662 ms, MAE od 10,1952 do 12, 1044 ms i CC od 0,8276 do 0,8833. Modeli trajanja vokala ostvaruju RMSE od 16,1226 do 19,9038 ms, MAE od 12,2446 do 15,3706 ms i CC od 0,8412 do 0,899. Kao i kod modela trajanja glasova najbolje performanse ima model trajanja konsonanata razvijen primenom aditivne regresije, ali u ovom slučaju bolji rezultati su dobijeni ukoliko se kao osnovni algoritam koristi M5PR. AR REPTree model trajanja vokala poseduje najbolje prediktivne performanse. Primena linearne regresije prilikom razvoja modela trajanja kako konsonanata, tako i vokala takođe daje najlošije rezultate. U cilju poboljšanja ostvarenih performansi razvijenih modela izvršeno je uklanjanje iz govorne baze onih glasova koji u najvećoj meri doprinose grešci predikcije (engl. *outliers*). Novodobijeni opseg trajanja glasova sadrži 96,27% ukupne količine podataka iz govorne baze. Ovaj skup je dobijen uzimajući u obzir raspodelu trajanja u bazi i broj pojavljivanja glasova koji imaju izuzetno male ili izuzetno velike vrednosti trajanja, odnosno nalaze se u blizini graničnih vrednosti opsega, t.j. oko 2 i 290 ms (slika 7.10). Raspodela trajanja glasova nakon uklanjanja glasova koji se nalaze u blizini granica opsega prikazana je na slici 7.11. Raspodela trajanja glasova u govornoj bazi približno odgovara gama raspodeli. Performanse modela trajanja glasova nakon uklanjanja glasova u blizini granica opsega prikazane su u tabeli 7.4.

Model trajanja konsonanata	**RMSE [ms]**	**MAE [ms]**	**CC**
LR	16,5562	12,1044	0,8276
M5P	14,0769	10,3277	0,8788
M5PR	14,4826	10,5714	0,8712
REPTree	15,1510	10,9402	0,8580
REPTree no prunning	15,0603	10,8792	0,8598
AR M5PR	13,8348	10,1952	0,8833
AR REPTree	13,9699	10,2276	0,8707
BAG M5PR	14,1849	10,3873	0,8770
BAG REPTree	14,9614	10,8181	0,8618
STACK M5P	14,0370	10,2833	0,8795

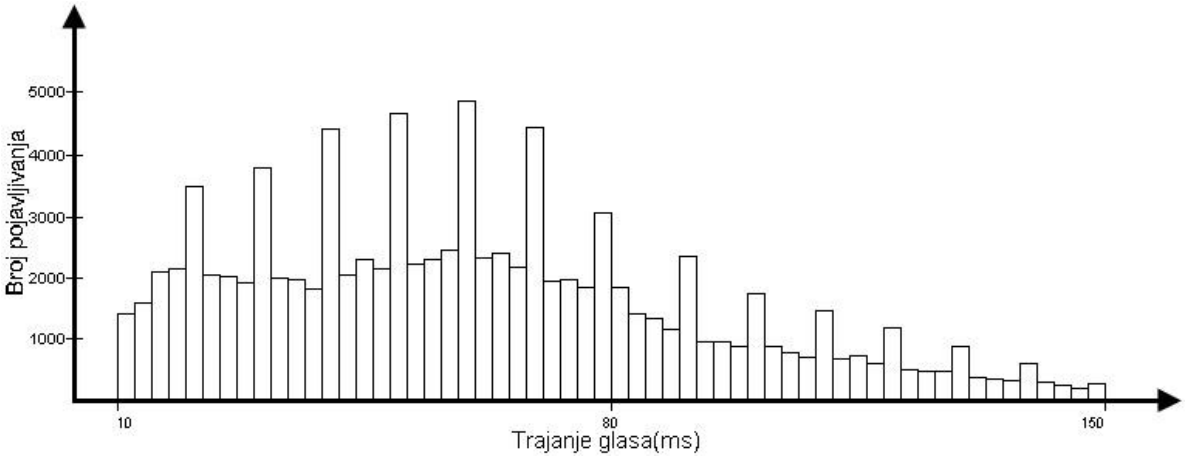
Tabela 7.2 – Prediktivne performanse modela trajanja konsonanata

Model trajanja vokala	RMSE [ms]	MAE [ms]	CC
LR	19,9038	15,3706	0,8412
M5P	16,6933	12,7239	0,8913
M5PR	17,2721	13,1076	0,8844
REPTree	17,8045	13,4284	0,8752
REPTree no prunning	17,6018	13,2887	0,8783
AR M5PR	16,5193	12,5885	0,8938
AR REPTree	16,1226	12,2446	0,8990
BAG M5PR	16,8572	12,7740	0,8891
BAG REPTree	17,5142	13,2146	0,8795
STACK M5P	16,6730	12,6544	0,8916

Tabela 7.3 – Prediktivne performanse modela trajanja vokala



Slika 7.10 – Raspodela trajanja glasova pre uklanjanja glasova u blizini granica opsega

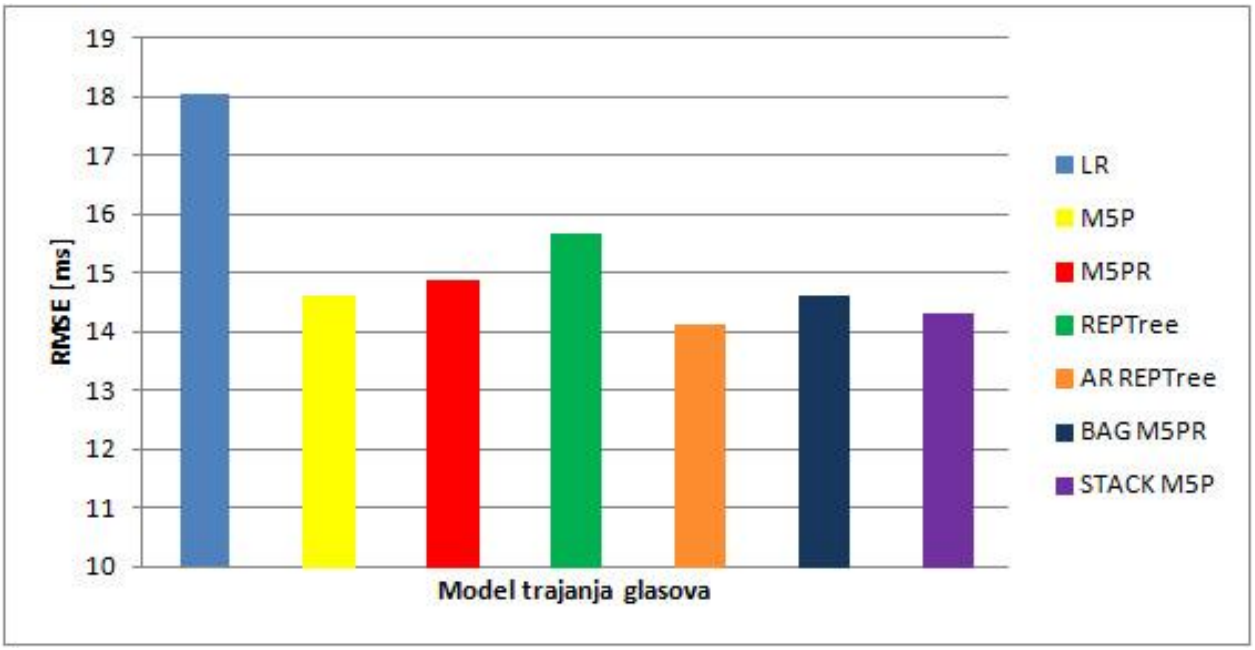


Slika 7.11 – Raspodela trajanja glasova nakon uklanjanja glasova u blizini granica opsega

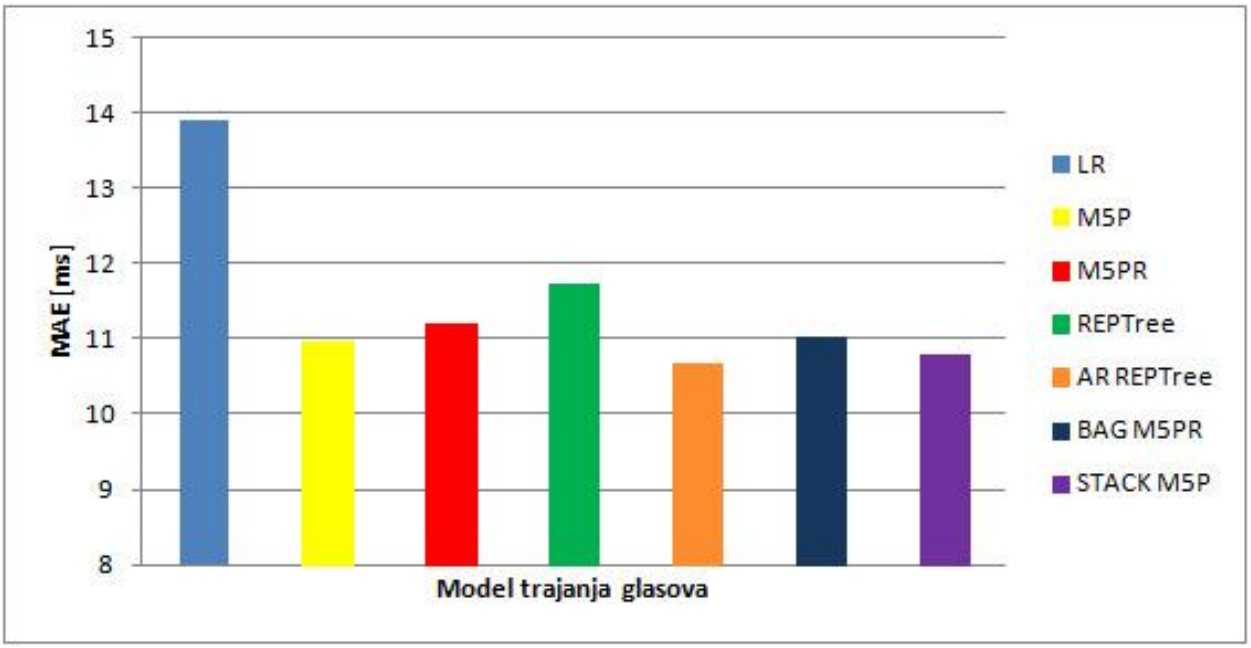
Model trajanja glasova	RMSE [ms]	MAE [ms]	CC
LR	18,0529	13,9174	0,8170
M5P	14,6028	10,9763	0,8845
M5PR	14,8914	11,1947	0,8796
REPTree	15,6526	11,7379	0,8660
AR REPTree	14,1246	10,6840	0,8924
BAG M5PR	14,6211	11,0218	0,8843
STACK M5P	14,3118	10,7948	0,8894

Tabela 7.4 – Prediktivne performanse modela trajanja glasova nakon uklanjanja glasova u blizini granica opsega

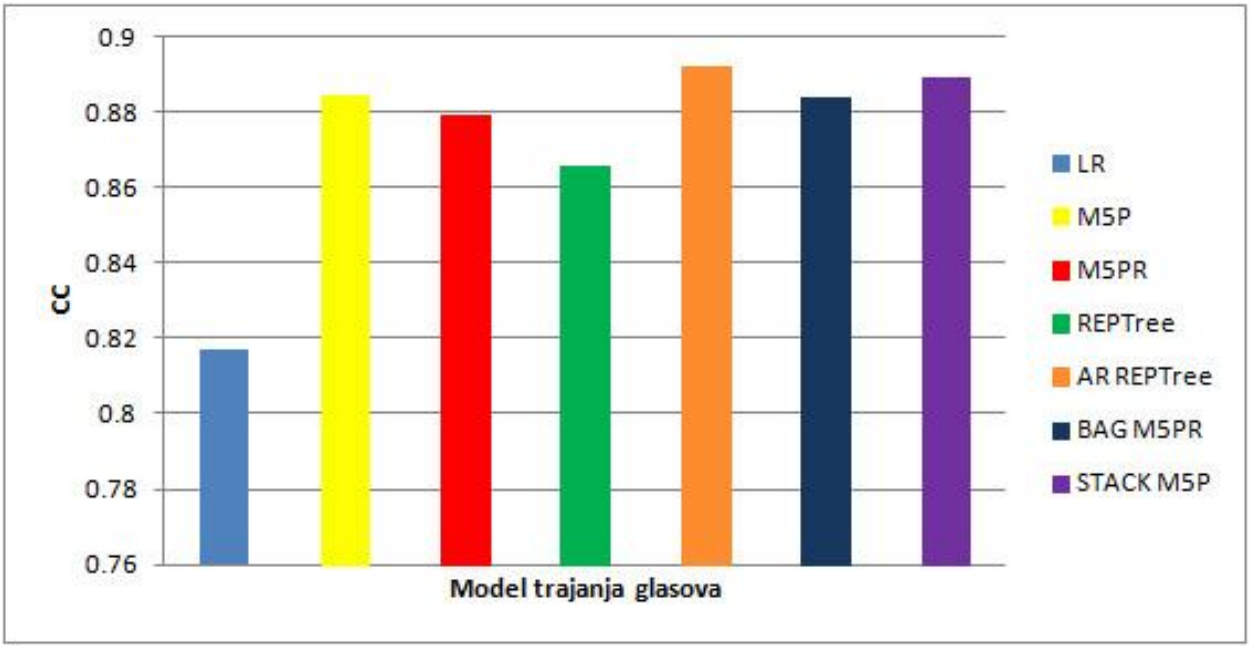
Kvantitativni pokazatelji RMSE, MAE i CC dobijenih modela trajanja glasova histogramom su prikazani na slikama 7.12, 7.13 i 7.14. Na osnovu dobijenih rezultata i odgovarajućih histograma može se uvideti da se AR REPTree model trajanja glasova i ovog puta pokazao kao model najboljih prediktivnih performansi. Primena linearne regresije ponovo daje model najlošijih performansi.



Slika 7.12 – RMSE kod različitih modela trajanja glasova nakon uklanjanja glasova u blizini granica opsega



Slika 7.13 – MAE kod različitih modela trajanja glasova nakon uklanjanja glasova u blizini granica opsega



Slika 7.14 – Koeficijent korelacije kod različitih modela trajanja glasova nakon uklanjanja glasova u blizini granica opsega

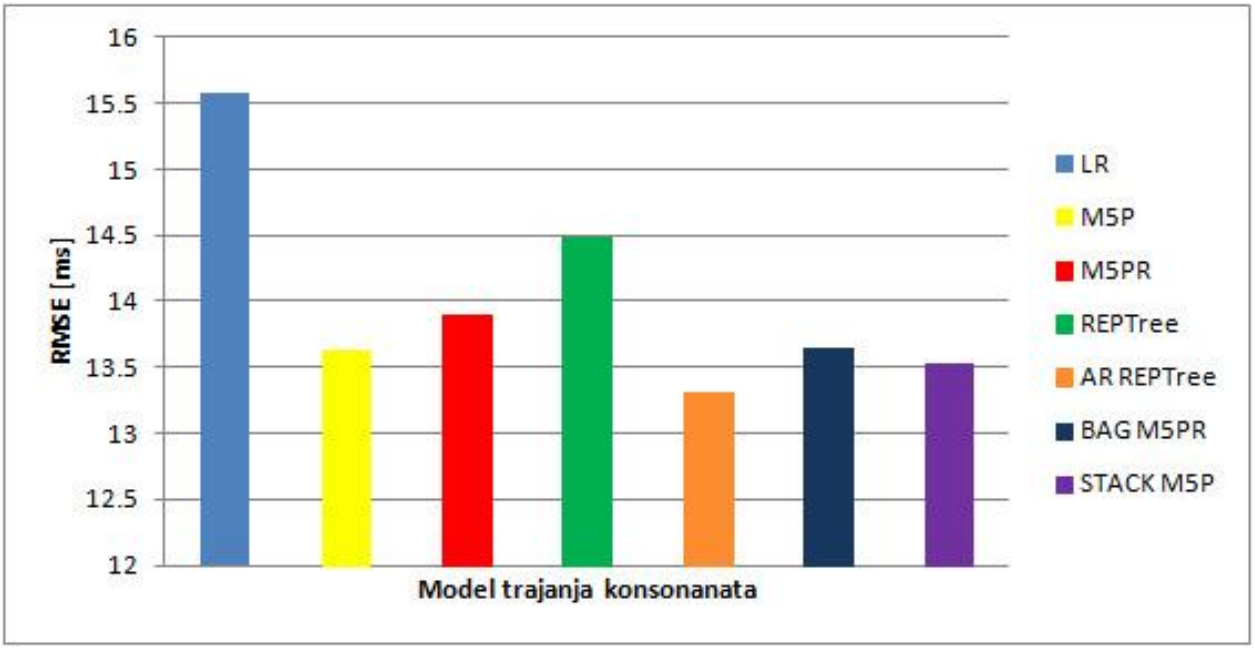
U tabelama 7.5 i 7.6 dati su RMSE, MAE i CC razvijenih modela trajanja konsonanata dobijeni nakon uklanjanja konsonanata koji se u govornoj bazi nalaze u blizini granica opsega, odnosno modeli trajanja vokala koji su dobijeni uklanjanjem iz govorne baze vokala u blizini granica opsega.

Model trajanja kosonanata	RMSE [ms]	MAE [ms]	CC
LR	15,5774	11,5999	0,8257
M5P	13,6274	10,0956	0,8698
M5PR	13,8944	10,2746	0,8643
REPTree	14,4790	10,6440	0,8516
AR REPTree	13,3149	9,9282	0,8762
BAG M5PR	13,6516	10,1497	0,8695
STACK M5P	13,5288	10,0532	0,8718

Tabela 7.5 – Prediktivne performanse modela trajanja konsonanata nakon uklanjanja konsonanata u blizini granica opsega

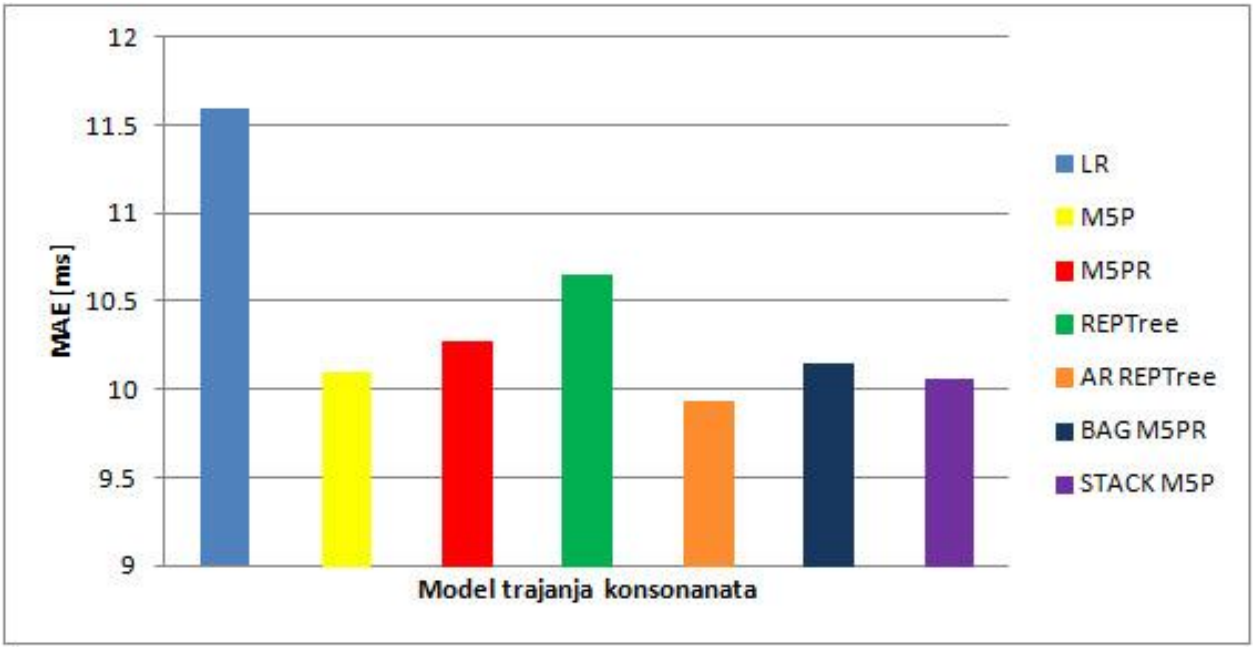
Model trajanja vokala	RMSE [ms]	MAE [ms]	CC
LR	17,0795	13,5574	0,8223
M5P	15,7466	12,1199	0,8710
M5PR	16,2816	12,5345	0,8615
REPTree	16,4006	12, 5757	0,8374
AR REPTree	15,4433	11,8954	0,8763
BAG M5PR	16,0174	12,3217	0,8666
STACK M5P	15,7223	12,0820	0,8714

Tabela 7.6 – Prediktivne performanse modela trajanja vokala nakon uklanjanja vokala u blizini granica opsega

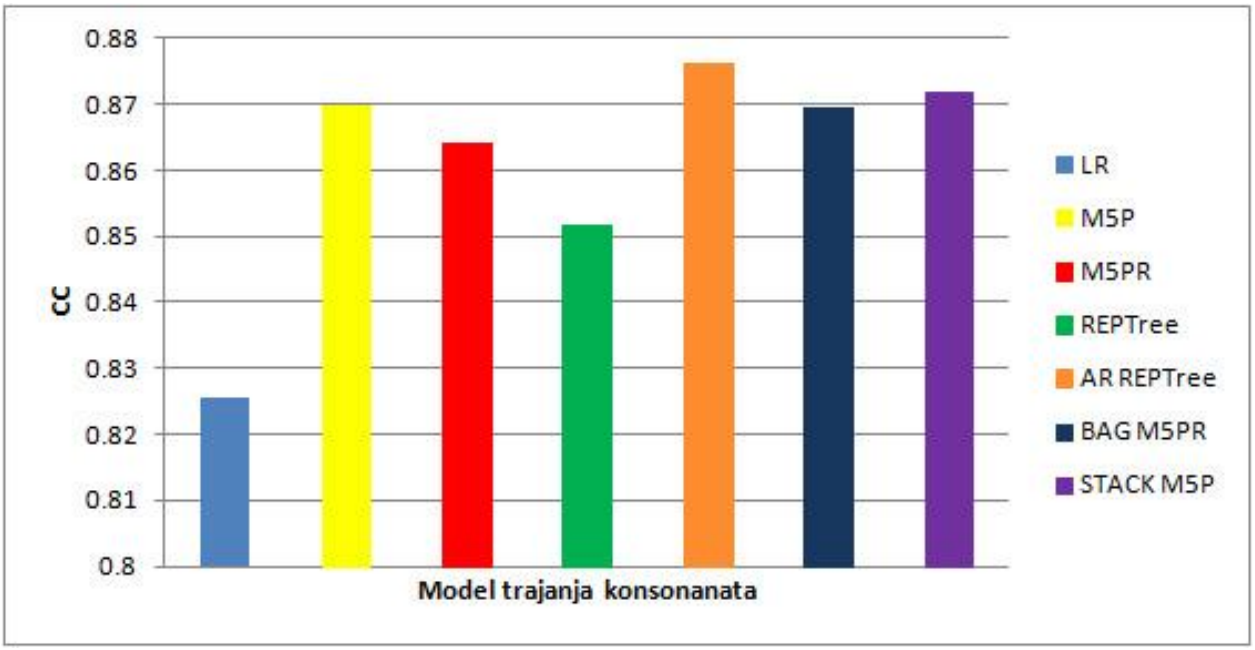


Slika 7.15 – RMSE kod različitih modela trajanja konsonanata nakon uklanjanja konsonanata u blizini granica opsega

Na slikama 7.15, 7.16, 7.17, 7.18, 7.19 i 7.20 histogramom su prikazani kvantitativni pokazatelji modela trajanja koji su dobijeni uklanjanjem iz govorne baze konsonanata, odnosno vokala koji se nalaze u blizini granica opsega.

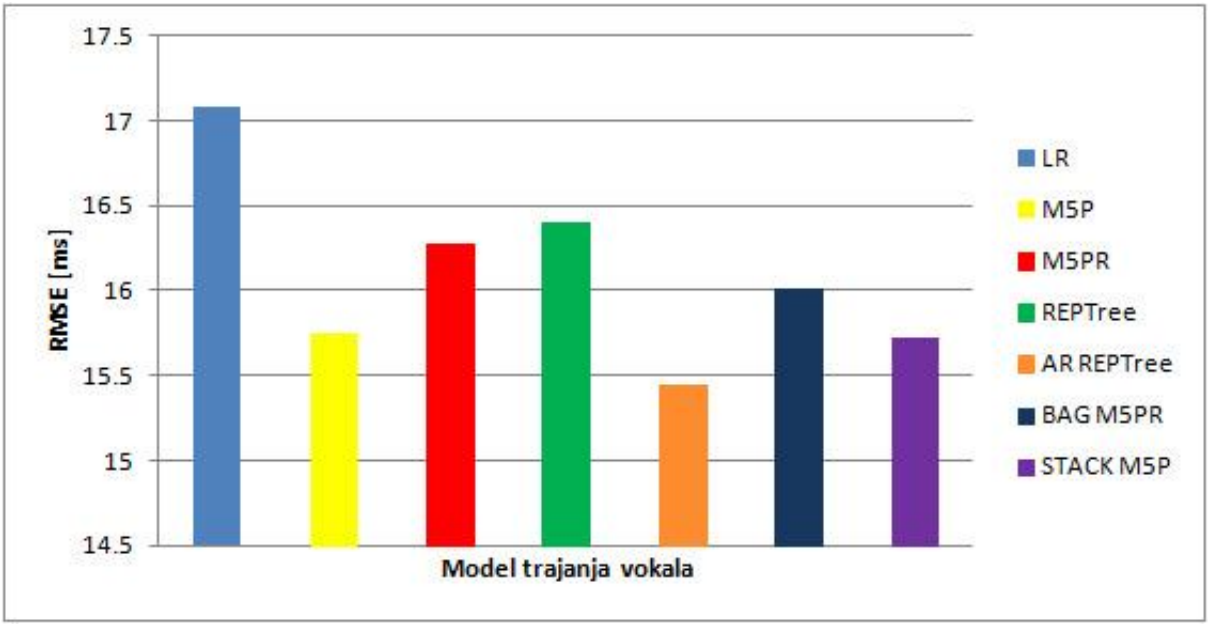


Slika 7.16 – MAE kod različitih modela trajanja konsonanata nakon uklanjanja konsonanata u blizini granica opsega

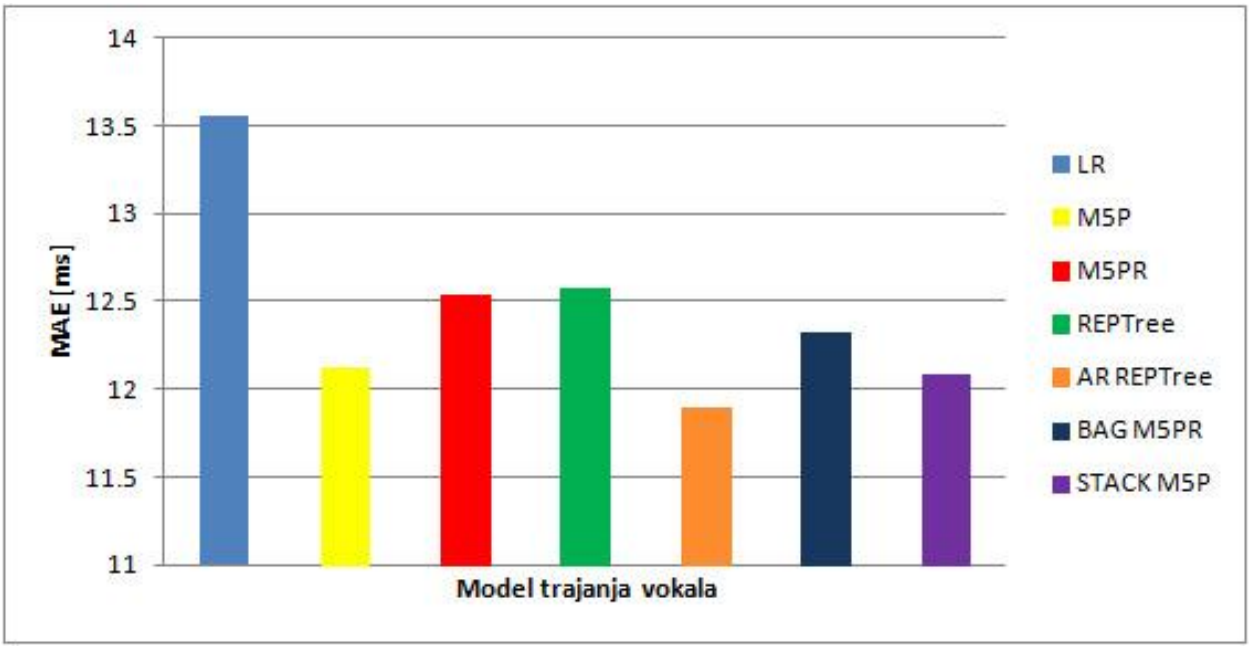


Slika 7.17 – Koeficijent korelacije kod različitih modela trajanja konsonanata nakon uklanjanja konsonanata u blizini granica opsega

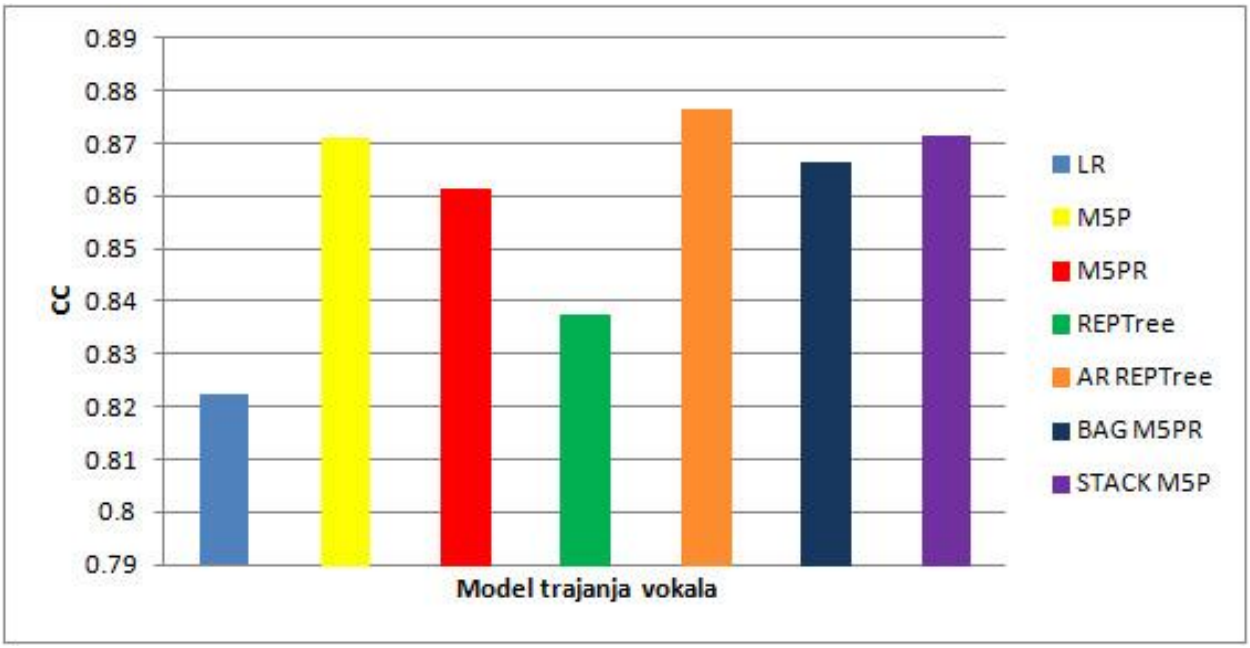
Na pomenutim slikama može se uočiti da primena AR REPTree algoritma u oba slučaja daje model trajanja najboljih prediktivnih performansi, dok primena linearne regresije daje modele najlošijih performansi.



Slika 7.18 – RMSE kod različitih modela trajanja vokala nakon uklanjanja vokala u blizini granica opsega



Slika 7.19 – MAE kod različitih modela trajanja vokala nakon uklanjanja vokala u blizini granica opsega



Slika 7.20 – Koeficijent korelacije kod različitih modela trajanja vokala nakon uklanjanja vokala u blizini granica opsega

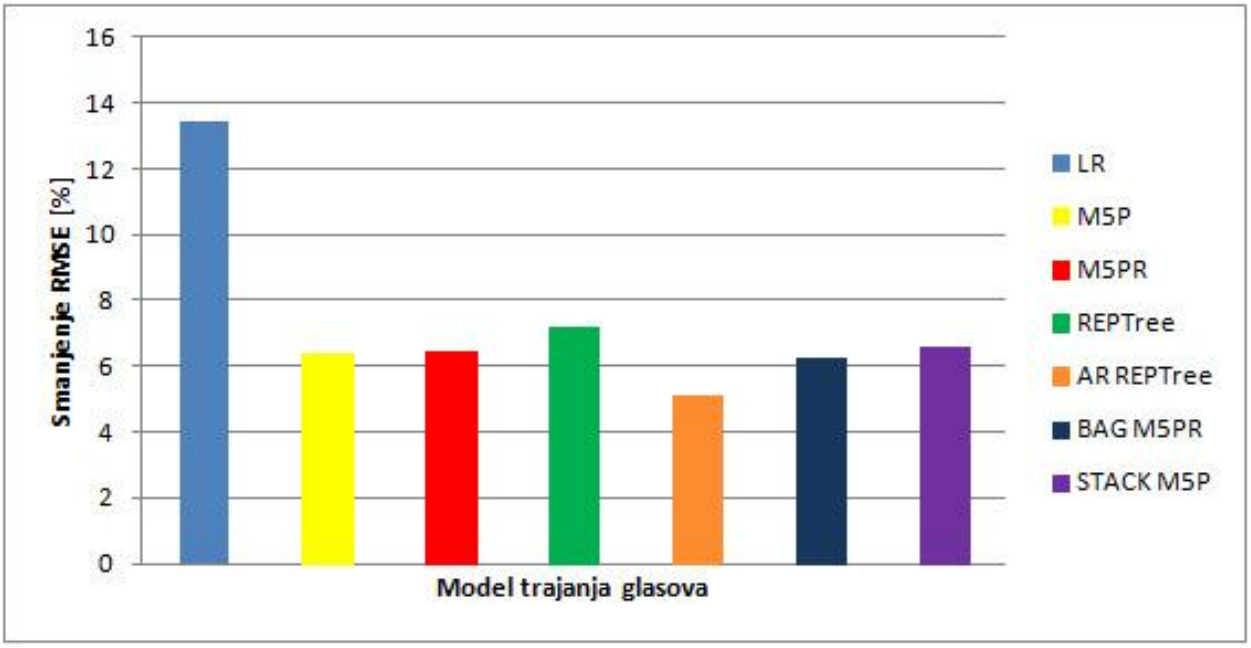
Procenat smanjenja RMSE razvijenih modela dobijenih nakon uklanjanja iz govorne baze glasova, odnosno konsonanata i vokala koji se nalaze u blizini granica opsega dat je u tabeli 7.7.

Model trajanja	Glasovi	Konsonanti	Vokali
LR	13,44%	5,91%	14,19%
M5P	6,39%	3,19%	5,67%
M5PR	6,43%	3,12%	5,50%
REPTree	7,22%	4,43%	7,88%
AR REPTree	5,13%	4,69%	4,21%
BAG M5PR	6,28%	3,76%	4,98%
STACK M5P	6,56%	3,62%	5,70%

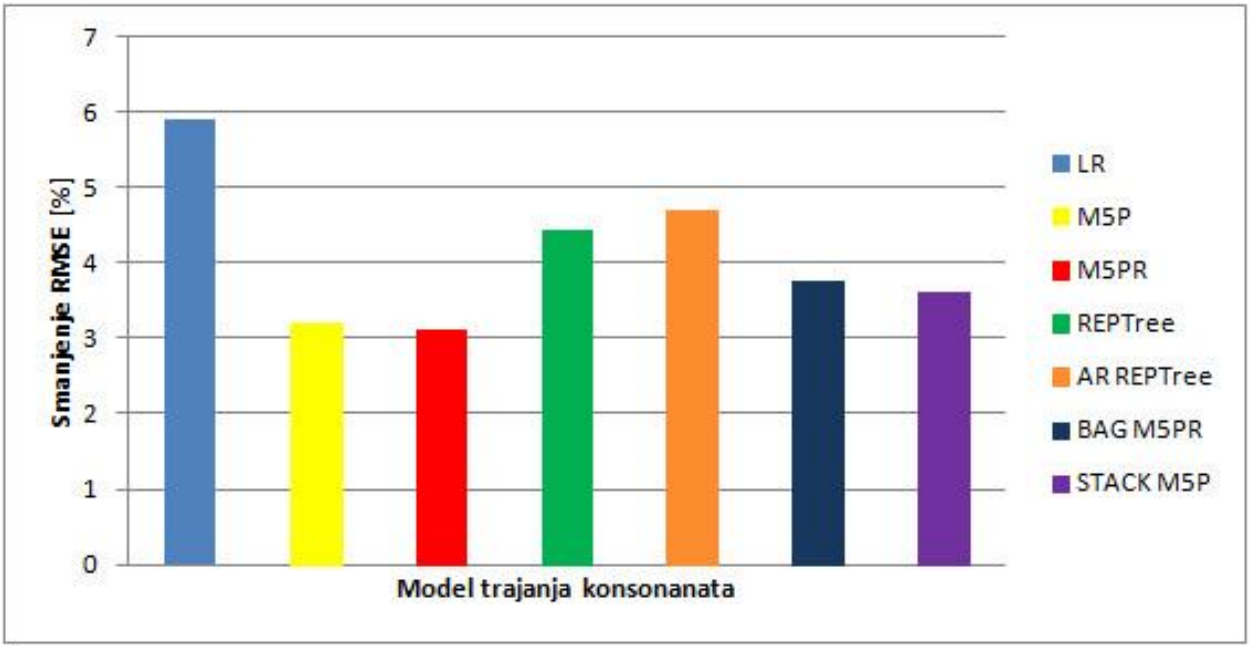
Tabela 7.7 – Smanjenje RMSE nakon uklanjanja glasova, konsonanata ili vokala u blizini granica opsega

Pomenuti procenat smanjenja RMSE prikazan je takođe histogramom na slikama 7.21, 7.22 i 7.23. Na osnovu prikazanih rezultata može se zapaziti da je smanjenje RMSE kod modela trajanja glasova u opsegu od 5,13% do 13,44%, kod modela trajanja konsonanata od 3,12% do 5,91% a kod modela trajanja vokala u opsegu od 4,21% do 14,19%. Takođe, može se uočiti da je procenat smanjenja RMSE najmanji kod AR REPTree modela trajanja glasova i AR REPTree modela trajanja vokala koji ujedno predstavljaju modele najboljih prediktivnih performansi. M5PR model trajanja konsonanata ostvaruje najmanji procenat smanjenja RMSE nakon uklanjanja konsonanata iz govorne baze koji se nalaze u blizini granica opsega. S druge strane,

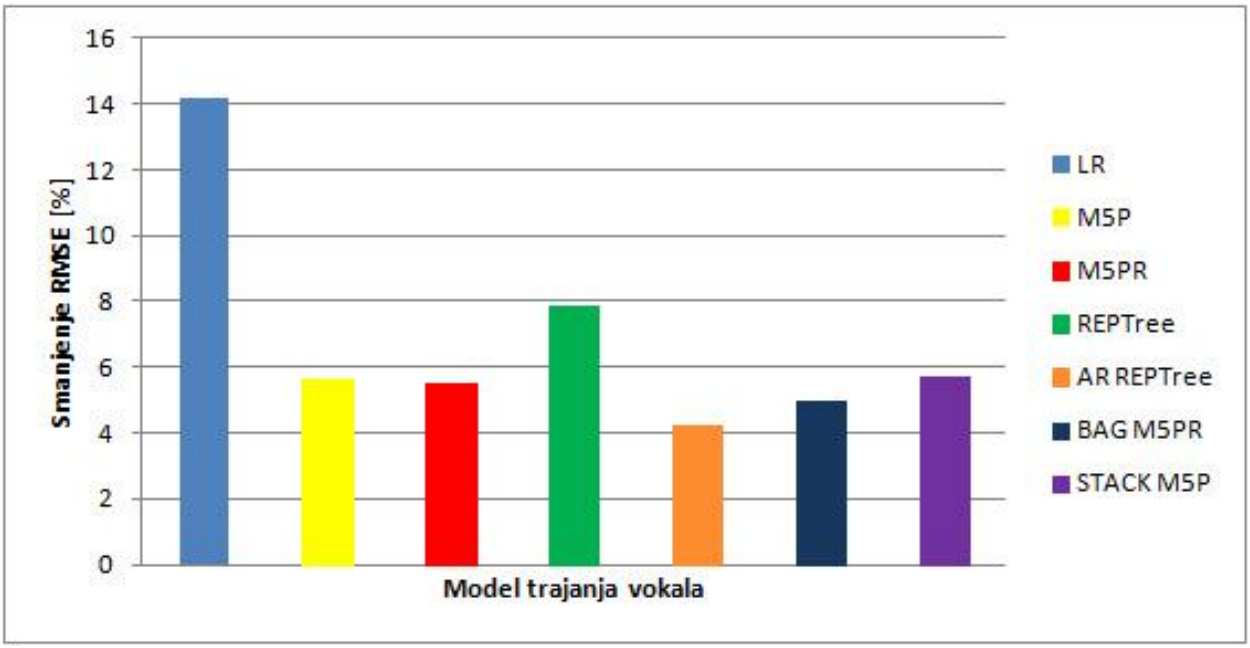
LR model koji predstavlja model najlošijih performansi postiže najveći procenat smanjenja RMSE nakon uklanjanja iz govorne baze glasova koji se nalaze u blizini granica opsega, a takođe i nakon uklanjanja konsonanata i vokala u blizini granica opsega.



Slika 7.21 – Smanjenje RMSE kod različitih modela trajanja glasova nakon uklanjanja glasova u blizini granica opsega



Slika 7.22 – Smanjenje RMSE kod različitih modela trajanja konsonanata nakon uklanjanja konsonanata u blizini granica opsega



Slika 7.23 – Smanjenje RMSE kod različitih modela trajanja vokala nakon uklanjanja vokala u blizini granica opsega

Prediktivne performanse modela zasnovanih na stablima odluke koji su razvijeni za predikciju trajanja glasova u turskom (Öztürk, 2005), češkom (Batušek, 2002), korejskom (Lee & Oh, 1999), srpskohrvatskom (Sečujski et al., 2011), hindu (Krishna & Murthy, 2004) i telugu jeziku (Krishna & Murthy, 2004) prikazani su u tabeli 7.8. Na osnovu kvantitativnih pokazatelja datih u tabeli 7.8 može se uočiti da rezultati ostvareni primenom regresionih stabala za razvoj modela trajanja glasova u srpskom jeziku RMSE 14,8914 ms i CC 0,8796 su uporedljivi ili čak prevazilaze rezultate koji se navode u literaturi a odnose se na druge jezike i prethodno razvijene modele trajanja.

Jezik	**RMSE** [ms]	**CC**
turski	20,04	0,78
češki	20,30	0,79
korejski	20,45	0,84
srpskohrvatski	15,85	0,91
hindu	27,14	0,75
telugu	22,86	0,80

Tabela 7.8 – Prediktivne performanse modela trajanja glasova u različitim jezicima

U tabelama 7.9 i 7.10 dati su rezultati dobijeni primenom regresionih stabala za razvoj modela trajanja konsonanata, odnosno vokala u litvanskom (Norkevičius & Raškinis, 2008) grčkom (Lazaridis et al., 2011) i engleskom jeziku (Lazaridis et al., 2011). Na osnovu prikazanih

rezultata može se zapaziti da RMSE 13,8947 ms ostvaren kod modela trajanja konsonanata i RMSE 16,2816 ms postignut kod modela trajanja vokala u srpskom jeziku su takođe uporedljivi ili čak prevazilaze rezultate dobijene za litvanski, grčki i engleski jezik.

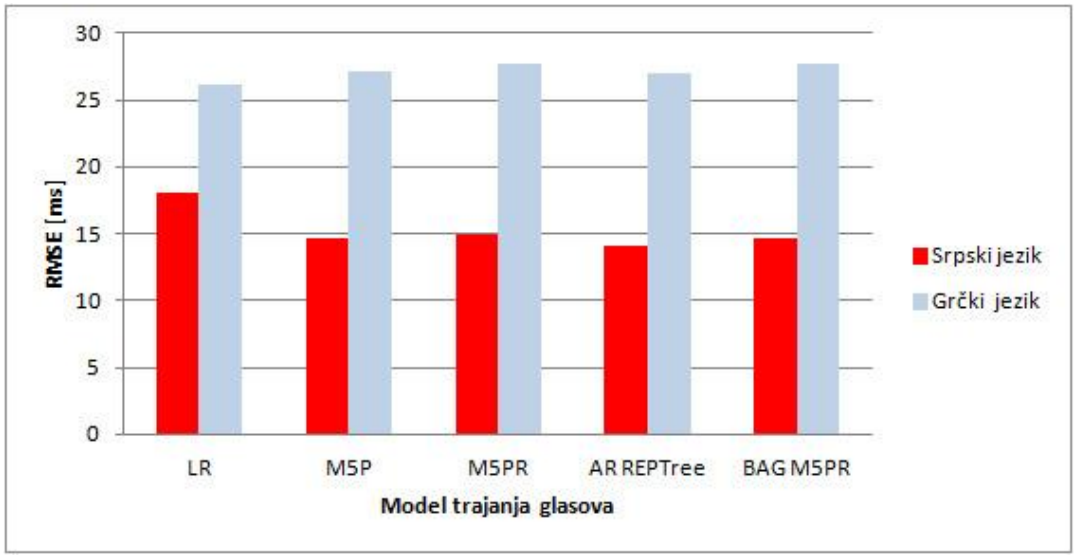
Jezik	RMSE [ms]
litvanski	16,70
grčki	29,13
engleski	20,74

Tabela 7.9 – Prediktivne performanse modela trajanja konsonanata u različitim jezicima

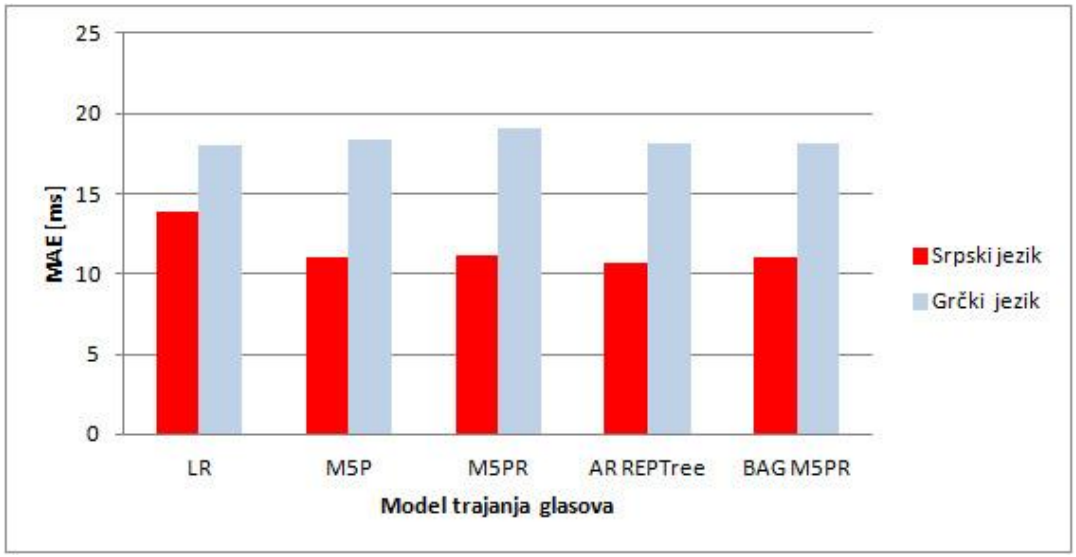
Jezik	RMSE [ms]
litvanski	18,30
grčki	26,04
engleski	25,46

Tabela 7.10 – Prediktivne performanse modela trajanja vokala u različitim jezicima

Na slikama 7.24 i 7.25 dat je uporedni prikaz prediktivnih performansi različitih modela koji su razvijeni za srpski, odnosno grčki jezik (Lazaridis et al., 2011). Na osnovu rezultata prikazanih na slici može se uočiti da modeli razvijeni za srpski jezik poseduju bolje performanse. Za razliku od srpskog jezika kod grčkog jezika model dobijen primenom linearne regresije poseduje najbolje performanse, dok je M5PR model najlošijih performansi.

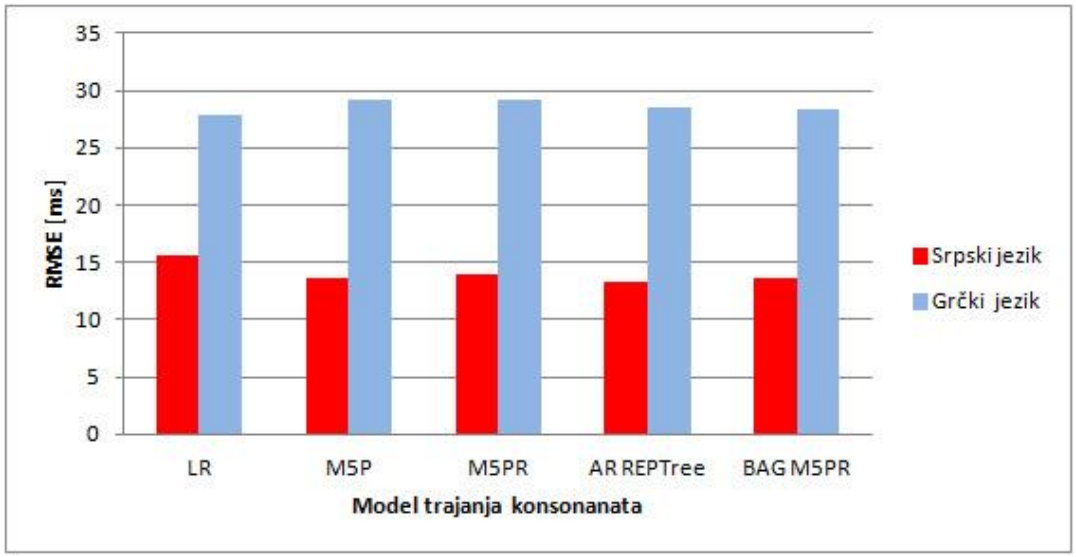


Slika 7.24 – Uporedni prikaz RMSE kod različitih modela trajanja glasova u srpskom i grčkom jeziku

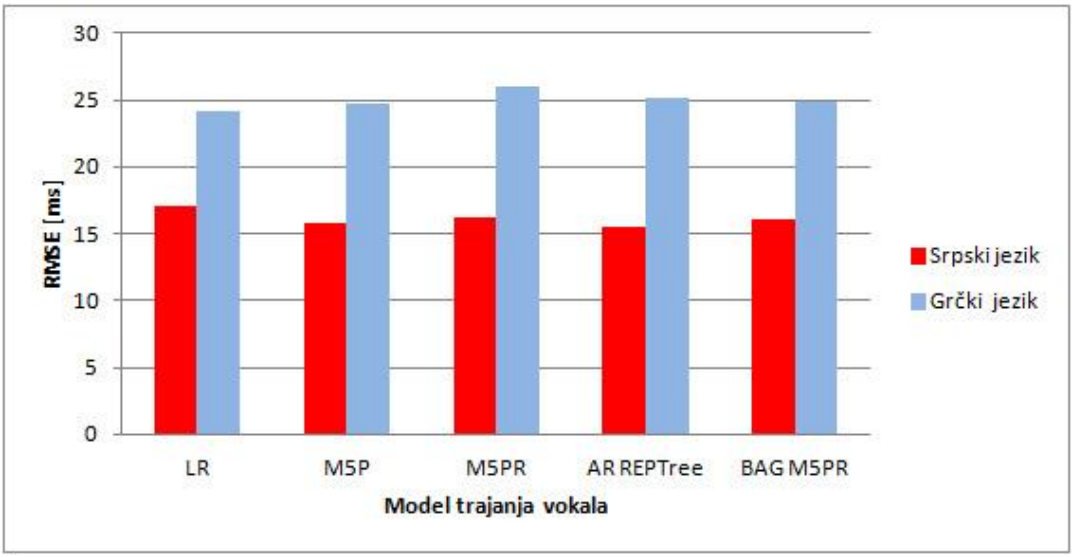


Slika 7.25 – Uporedni prikaz MAE kod različitih modela trajanja glasova u srpskom i grčkom jeziku

Na slikama 7.26 i 7.27 dat je uporedni prikaz kvantitativnog pokazatelja RMSE koji je postignut kod različitih modela trajanja konsonanata i vokala razvijenih za srpski, odnosno grčki jezik (Lazaridis et al., 2011). Na osnovu prikazanih rezultata za predikciju trajanja konsonanata i vokala u grčkom jeziku može se zapaziti da najbolje performanse takođe poseduje LR model, dok je M5P najlošiji model za predikciju trajanja konsonanata a M5PR model daje najlošiju predikciju trajanja vokala u grčkom jeziku.

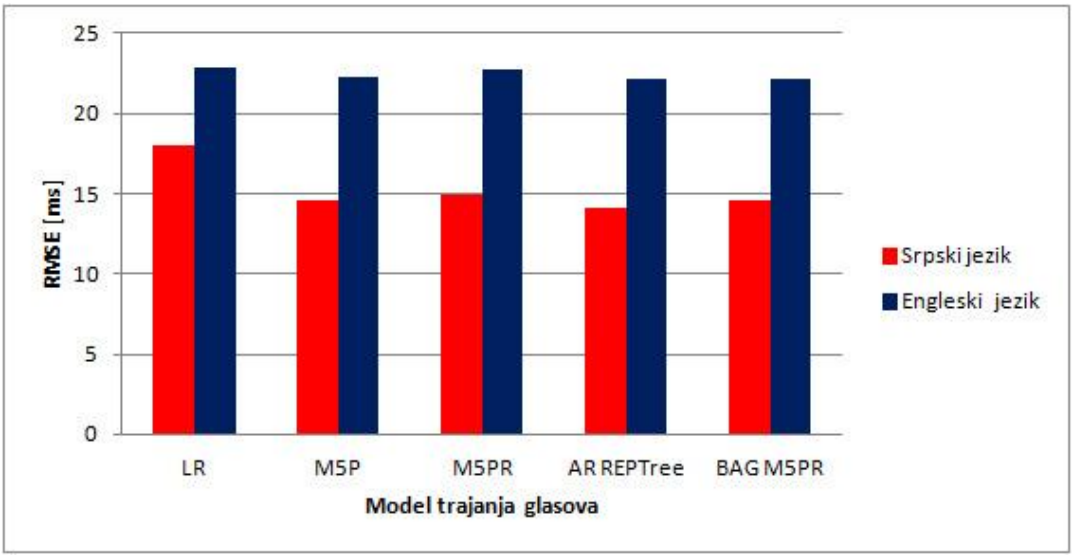


Slika 7.26 – Uporedni prikaz RMSE kod različitih modela trajanja konsonanata u srpskom i grčkom jeziku

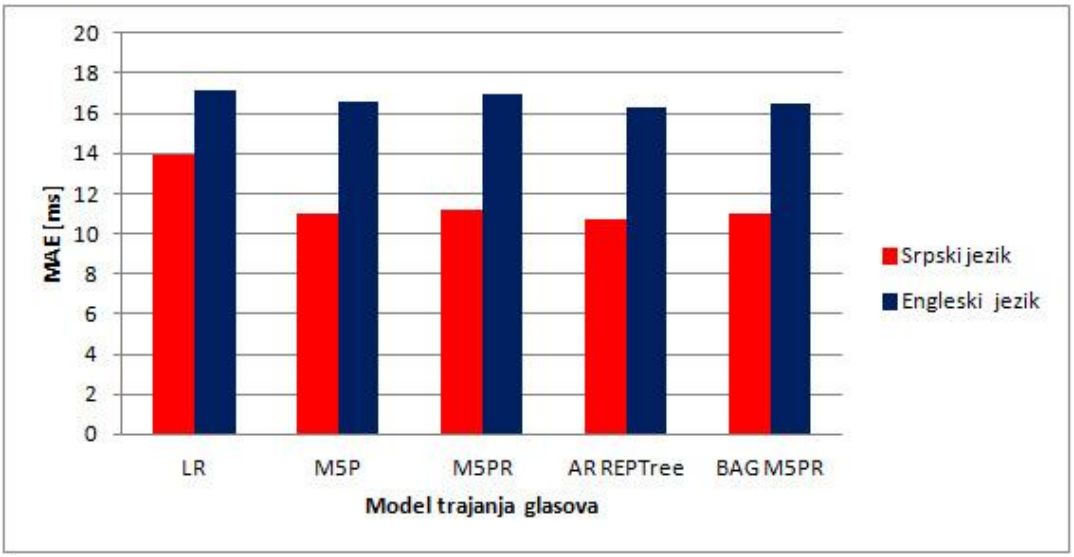


Slika 7.27 – Uporedni prikaz RMSE kod različitih modela trajanja vokala u srpskom i grčkom jeziku

Na slikama 7.28 i 7.29 dat je uporedni prikaz prediktivnih performansi različitih modela trajanja glasova koji su razvijeni za srpski, odnosno engleski jezik (Lazaridis et al., 2011). Na osnovu rezultata prikazanih na slikama može se uočiti da modeli razvijeni za srpski jezik poseduju bolje performanse. Kod engleskog jezika BAG M5PR model poseduje najbolje performanse, dok je najlošiji model, kao i kod srpskog jezika, dobijen primenom linearne regresije.

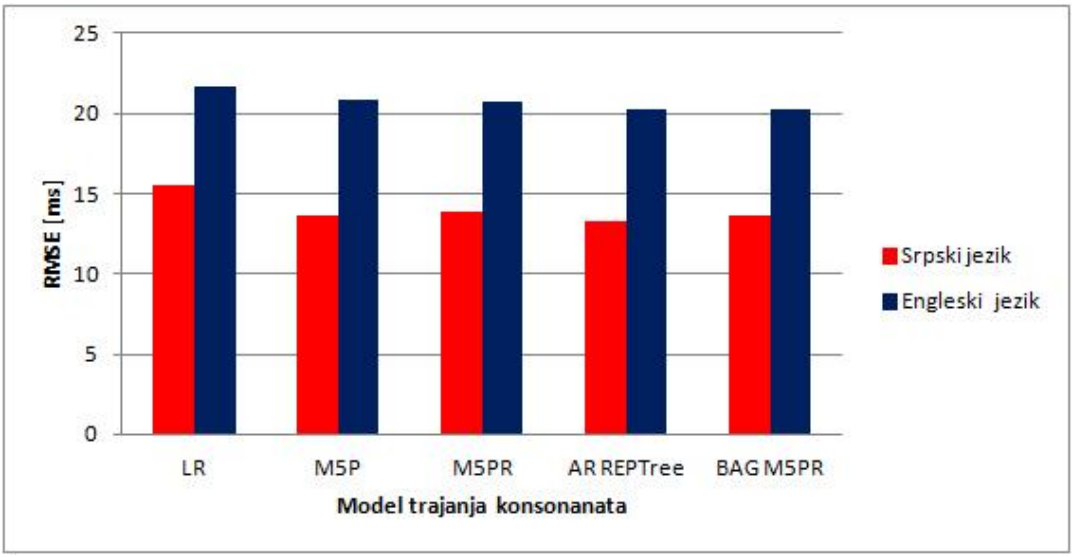


Slika 7.28 – Uporedni prikaz RMSE kod različitih modela trajanja glasova u srpskom i engleskom jeziku

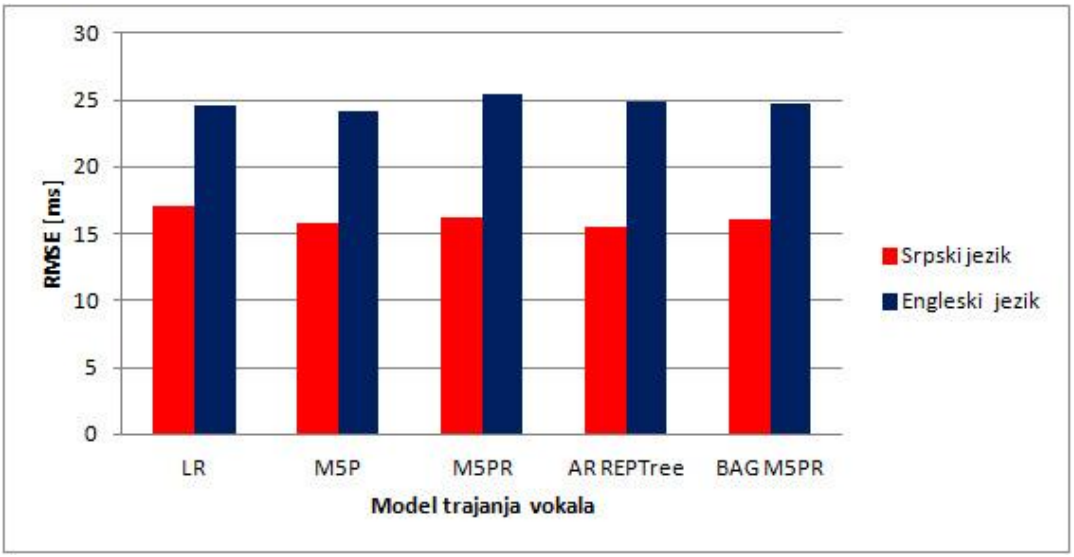


Slika 7.29 – Uporedni prikaz MAE kod različitih modela trajanja glasova u srpskom i engleskom jeziku

Na slikama 7.30 i 7.31 dat je uporedni prikaz kvantitativnog pokazatelja RMSE koji je postignut kod različitih modela trajanja konsonanata, odnosno vokala razvijenih za srpski, odnosno engleski jezik (Lazaridis et al., 2011). Na osnovu prikazanih rezultata za predikciju trajanja konsonanata u engleskom jeziku može se zapaziti da najbolje performanse poseduju AR REPTree i BAG M5PR modeli, dok je M5P najbolji model za predikciju trajanja vokala. Najlošiju predikciju trajanja konsonanata u engleskom jeziku daje LR model, dok M5PR model predstavlja najlošiji model trajanja vokala.



Slika 7.30 – Uporedni prikaz RMSE kod različitih modela trajanja konsonanata u srpskom i engleskom jeziku



Slika 7.31 – Uporedni prikaz RMSE kod različitih modela trajanja vokala u srpskom i enaleskom ieziku

Evaluacija razvijenih modela trajanja, zasnovanih na regresionim stablima, kao i poređenje dobijenih rezultata sa rezultatima AlfaNum (Sečujski et al., 2001) modela trajanja glasova izvršena je testiranjem modela na dve grupe test rečenica. U prvoj grupi nalazi se osam rečenica koje su učestvovale prilikom razvoja modela trajanja u fazi obučavanja modela, dok drugu grupu rečenica čini deset rečenica koje nisu učestvovale u obučavanju modela. Pomenute grupe rečenica navedene su u nastavku.

Rečenice koje su učestvovale u fazi obučavanja modela:

1. Ali bili su suviše spori.
2. Ali ipak ima nešto što liči na čudo.
3. Ali ipak je bilo strašno egzaktno.
4. Aplikacija je aktivna stalno.
5. Asimov je formulisao tri zakona robotike.
6. Atentat je izvršen pre nedelju dana.
7. Baronica i grof se zgledaše.
8. Beogradski autori su ponudili preko četrdeset radova.

Rečenice koje nisu učestvovale u fazi obučavanja modela:

1. Spor sam nastavila kao zakonski naslednik majke.
2. Ići ću do Vrhovnog suda.
3. Domaćínstva u Srbiji najčešće su četvoročlana.
4. Ne znam šta bih vam rekao.
5. Priroda nije uvek simetrična.

6. Probudila nas je dečja vika iz komšiluka.
7. Na odgovoru su insistirali poslanici radikala.
8. Poslanici Valonci napustili su salu u znak protesta.
9. Izgradnja traje bar deset godina.
10. Radni vek nuklearke je trideset godina.

U tabelama 7.11-7.16. prikazane su vrednosti greške predikcije, odnosno RMSE koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela za različite grupe fonema. Testiranje modela vršeno je na rečenicama koje su učestvovale u obuci modela. Model M5PR (1) predstavlja model trajanja glasova, dok M5PR (2) podrazumeva primenu modela trajanja vokala, odnosno konsonanata u zavisnosti od toga koji je fonem u pitanju. Kod AlfaNum modela primenjuje se poseban model trajanja za svaki od fonema. Na osnovu rezultata prikazanih u tabeli 7.11 može se uočiti da je greška predikcije trajanja nenaglašenih vokala najmanja u slučaju kada se primenjuje M5PR (2) model, odnosno model za predikciju trajanja vokala. Prilikom predikcije trajanja naglašenih vokala AlfaNum model pokazuje bolje performanse od preostala dva modela. Na osnovu rezultata prikazanih u tabeli 7.13 može se zapaziti da primena modela trajanja konsonanata pri predikciji trajanja ploziva daje najbolje rezultate. Greška predikcije trajanja frikativa najmanja je kod primene AlfaNum modela. Rezultati prikazani u tabeli 7.15 pokazuju da je greška predikcije trajanja afrikata najmanja ukoliko se prilikom predikcije primeni model trajanja konsonanata. Na osnovu rezultata prikazanih u tabeli 7.16 može se uočiti da AlfaNum model pokazuje neznatno bolje performanse priklikom predikcije trajanja sonanata od M5PR (1) modela.

Fonem	**M5PR(1) model**	**M5PR(2) model**	**AlfaNum model**
A	13,626	11,305	10,969
E	18,948	15,897	15,175
I	11,850	10,252	11,111
O	16,765	15,454	20,505
U	9,474	10,426	8,498
Y	6,694	11,059	6,278
Yv	2,786	6,897	2,241
Nenaglašeni vokali	14,015	12,660	13,297

Tabela 7.11 – Greška predikcije trajanja nenaglašenih vokala kod različitih modela

Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
A	42,610	30,802	14,596
E	15,455	15,940	12,418
I	19,125	21,426	29,955
O	30,823	30,961	34,688
U	19,860	18,799	18,766
Yv	9,086	22,097	22,008
Naglašeni vokali	28,962	25,092	23,964

Tabela 7.12 – Greška predikcije trajanja naglašenih vokala kod različitih modela

Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
Po	16,933	13,627	14,892
Pe	8,677	9,069	7,153
To	8,163	9,585	8,225
Te	7,197	7,778	7,011
Ko	12,978	9,758	30,775
Ke	8,378	6,905	14,264
Bo	27,116	8,266	8,021
Be	6,983	7,020	13,117
Do	13,912	14,482	14,161
De	8,564	8,405	5,777
Go	6,871	5,779	5,153
Ge	11,752	13,289	5,652
Plozivi	11,822	9,766	13,522

Tabela 7.13 – Greška predikcije trajanja ploziva kod različitih modela

Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
S	9,542	8,475	14,634
SH	23,167	22,474	13,910
F	18,214	6,482	8,481
Z	10,072	11,167	8,082
V	10,794	9,858	11,949
Frikativi	14,483	13,214	12,867

Tabela 7.14 – Greška predikcije trajanja frikativa kod različitih modela

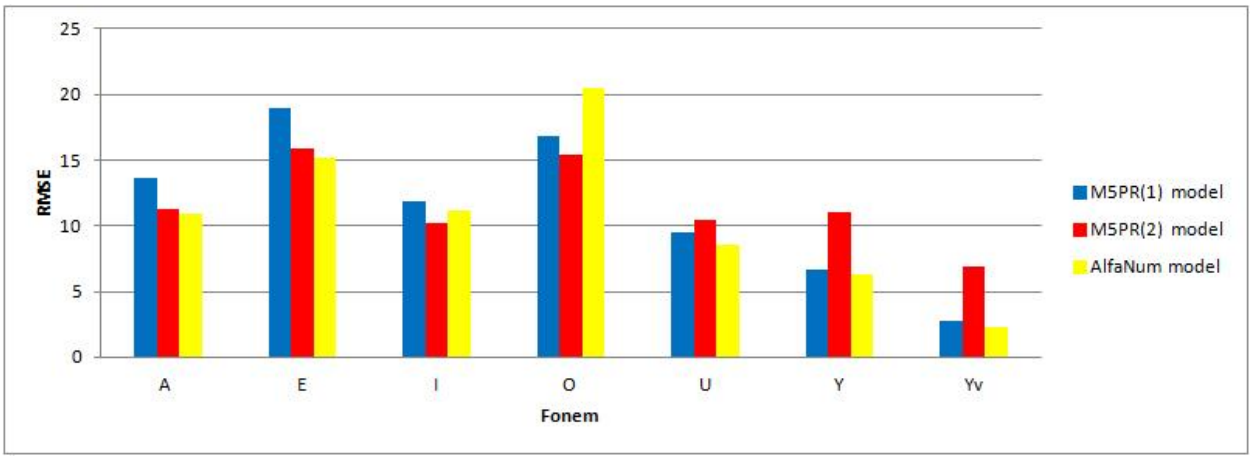
Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
Co	7,246	5,396	18,010
Ce	4,180	6,611	17,839
CHo	13,235	10,588	12,009
CHe	7,415	8,391	10,871
Afrikati	9,112	8,326	14,396

Tabela 7.15 – Greška predikcije trajanja afrikata kod različitih modela

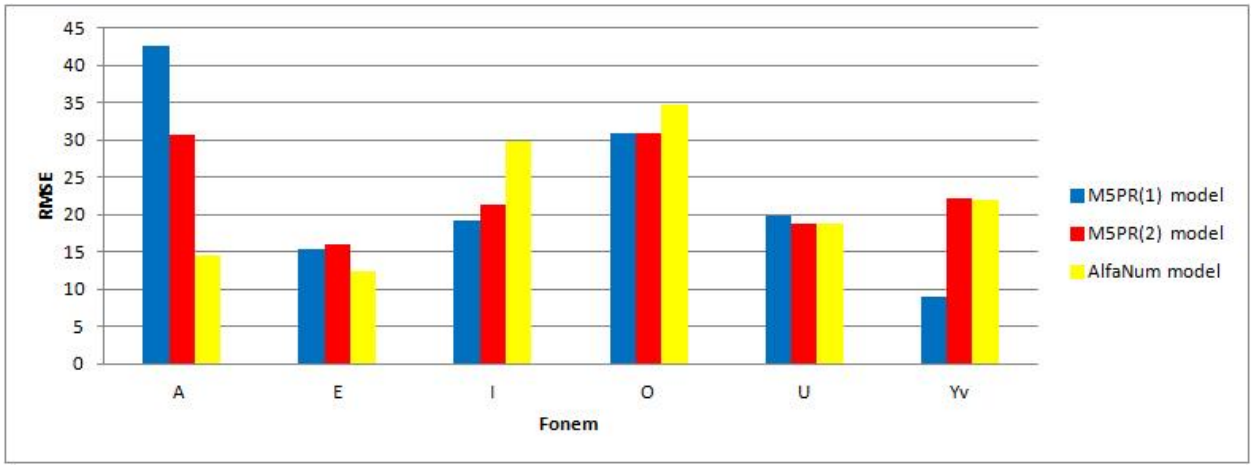
Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
M	8,742	8,086	9,445
N	15,470	14,066	14,012
L	7,257	11,263	8,210
LJ	4,222	4,347	2,496
R	4,909	5,160	4,578
J	16,117	17,408	18,347
Sonanti	10,88989	11,36486	10,861

Tabela 7.16 – Greška predikcije trajanja sonanata kod različitih modela

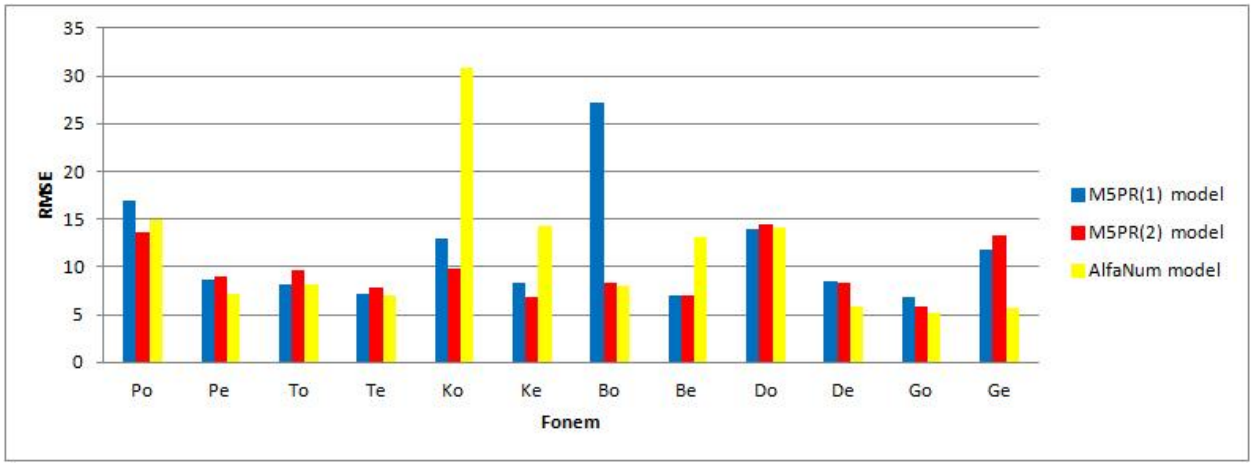
Na slikama 7.32 - 7.37 dat je uporedni prikaz vrednosti greške predikcije, odnosno RMSE koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela trajanja za različite grupe fonema. U okviru određene grupe fonema histogramom su prikazane vrednosti greške predikcije za svaki od fonema koji pripada datoj grupi a pojavljuje se u okviru test rečenica.



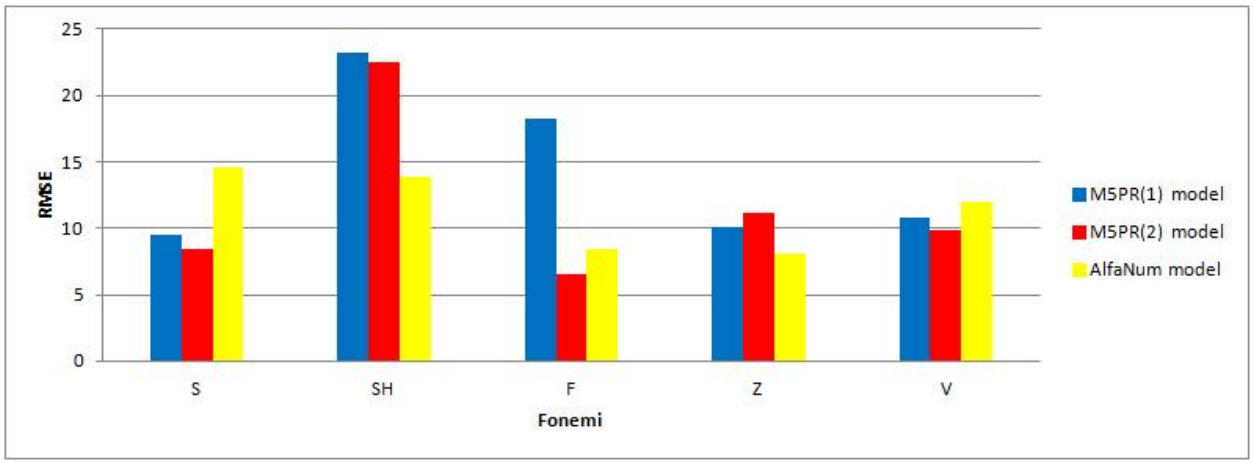
Slika 7.32 – Uporedni prikaz greške predikcije trajanja nenaglašenih vokala kod različitih modela trajanja



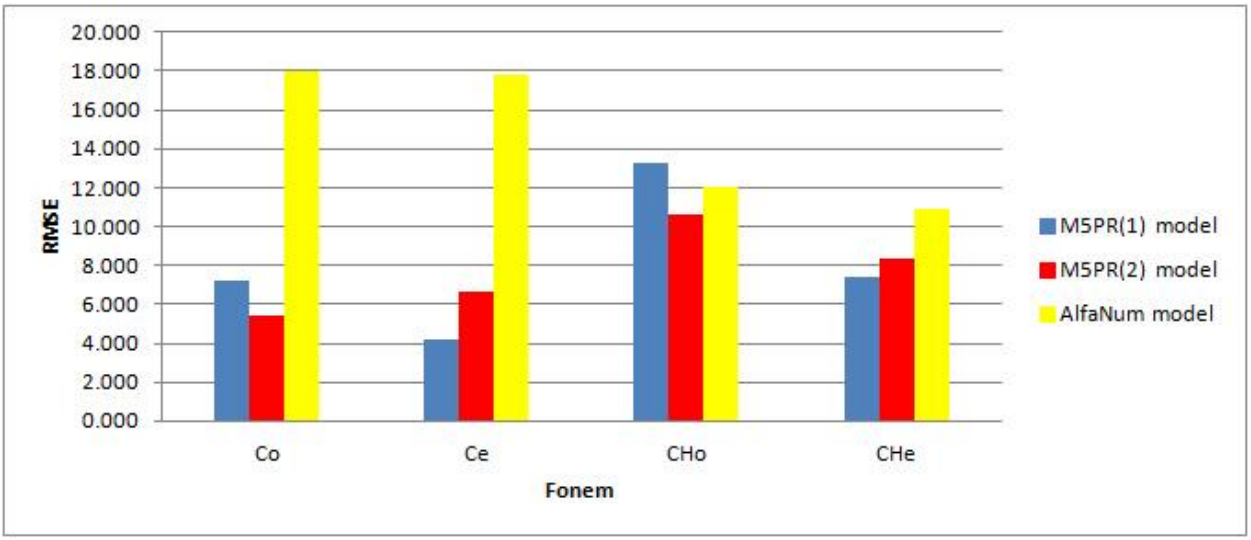
Slika 7.33 – Uporedni prikaz greške predikcije trajanja naglašenih vokala kod različitih modela trajanja



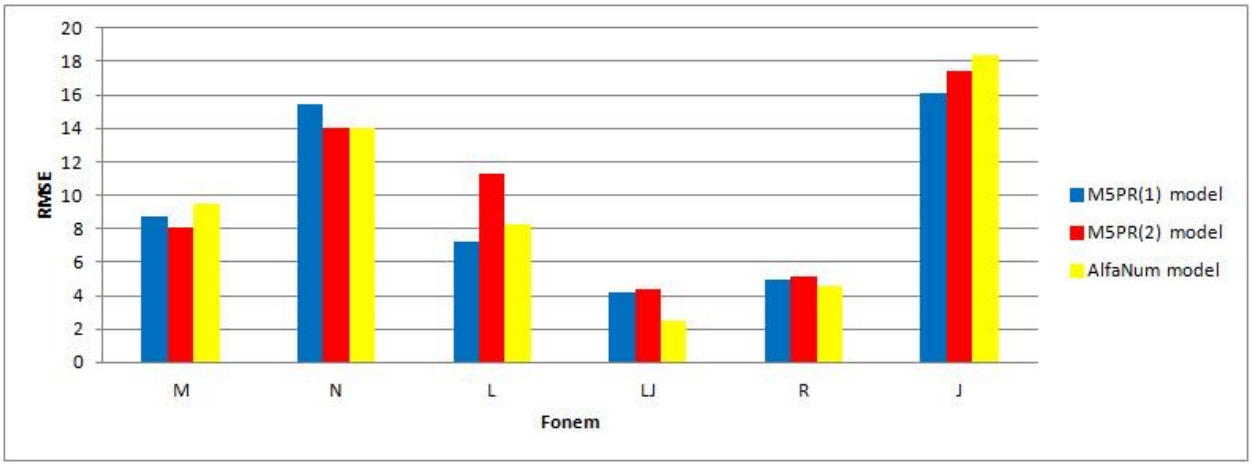
Slika 7.34 – Uporedni prikaz greške predikcije trajanja ploziva kod različitih modela trajanja



Slika 7.35 – Uporedni prikaz greške predikcije trajanja frikativa kod različitih modela trajanja

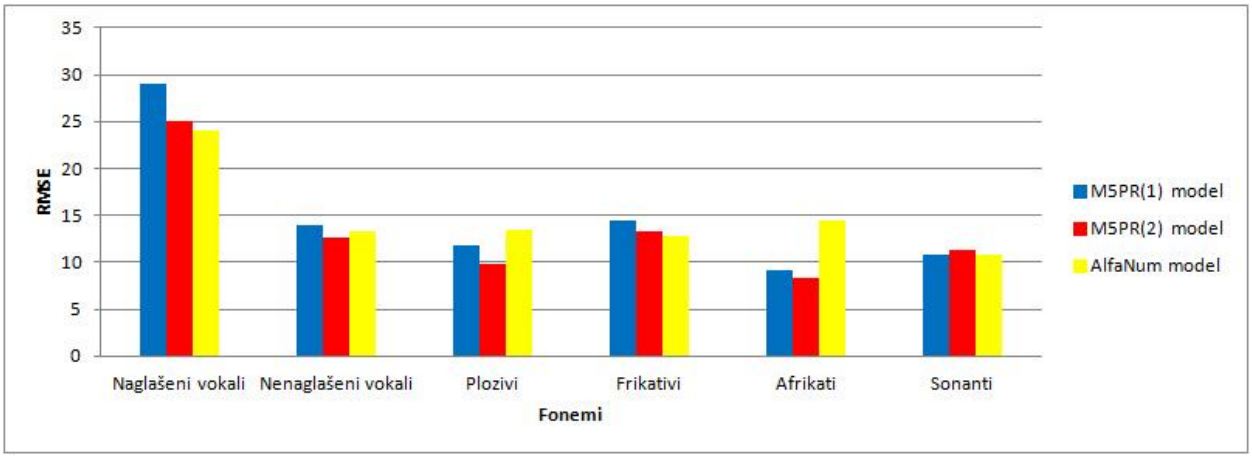


Slika 7.36 – Uporedni prikaz greške predikcije trajanja afrikata kod različitih modela trajanja



Slika 7.37 – Uporedni prikaz greške predikcije trajanja sonanata kod različitih modela trajanja

Na slici 7.38 dat je uporedni prikaz vrednosti RMSE koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela trajanja za svaku grupu fonema. Rezultati testiranja na rečenicama koje su učestvovale u fazi obučavanja modela pokazuju da su prediktivne performanse M5PR (2) modela bolje nego M5PR (1) modela, odnosno da je greška predikcije trajanja naglašenih i nenaglašenih vokala manja ukoliko se primeni model trajanja vokala prilikom predikcije trajanja. Takođe, greška predikcije je manja i ako se u slučaju predikcije trajanja određene grupe konsonanata primeni model trajanja konsonanata umesto modela trajanja glasova.



Slika 7.38 – Uporedni prikaz greške predikcije trajanja različitih grupa fonema kod različitih modela trajanja

U drugom testu testiranje modela trajanja vršeno je na rečenicama koje nisu učestvovale u fazi obuke modela. U tabelama 7.17 – 7.22 prikazane su vrednosti greške predikcije koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela za predikciju trajanja različitih grupa fonema. Na osnovu rezultata prikazanih u tabeli 7.17 može se uočiti da su performanse predikcije trajanja nenaglašenih vokala kod modela trajanja vokala (M5PR (2)) i AlfaNum modela gotovo identične.

Fonem	**M5PR(1) model**	**M5PR(2) model**	**AlfaNum model**
A	15,433	14,312	14,113
E	12,632	11,991	10,469
I	10,508	10,384	10,929
O	20,041	19,088	19,356
U	13,397	14,299	14,545
Y	16,338	14,611	14,830
Nenaglašeni vokali	14,630	13,895	13,870

Tabela 7.17 – Greška predikcije trajanja nenaglašenih vokala kod različitih modela

Prilikom predikcije trajanja naglašenih vokala AlfaNum model pokazuje najbolje performanse. Na osnovu rezultata prikazanih u tabeli 7.19 može se zapaziti da primena modela trajanja konsonanata pri predikciji trajanja ploziva daje najbolje rezultate. Greška predikcije trajanja frikativa najmanja je kod primene modela trajanja konsonanata (M5PR (2)).

Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
A	42,486	34,980	22,837
E	11,843	10,807	8,929
I	25,579	25,260	23,827
O	14,547	13,978	11,541
U	28,086	20,392	4,888
Yv	3,234	3,858	7,569
Naglašeni vokali	29,887	25,338	18,087

Tabela 7.18 – Greška predikcije trajanja naglašenih vokala kod različitih modela

Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
Po	4,319	4,806	4,709
Pe	7,690	7,893	6,119
To	15,139	13,993	14,613
Te	5,165	5,564	9,892
Ko	14,044	12,374	20,675
Ke	11,932	10,822	20,429
Bo	14,363	15,142	13,170
Be	3,794	3,772	4,061
Do	15,558	13,764	11,047
De	4,229	4,607	6,954
Go	12,877	12,515	14,966
Ge	7,312	4,579	7,949
Plozivi	11,215	10,342	13,620

Tabela 7.19 – Greška predikcije trajanja ploziva kod različitih modela

Fonem	M5PR(1) model	M5PR(2) model	AlfaNum model
S	17,641	16,670	18,328
SH	24,899	17,389	28,405
H	12,810	16,951	10,034
Z	9,669	9,567	6,744
V	12,061	13,245	12,213
Frikativi	16,068	15,236	16,622

Tabela 7.20 – Greška predikcije trajanja frikativa kod različitih modela

Rezultati prikazani u tabeli 7.21 pokazuju da je greška predikcije trajanja afrikata najmanja ukoliko se prilikom predikcije primeni model trajanja konsonanata. Na osnovu rezultata prikazanih u tabeli 7.22 može se uočiti da model M5PR (1) pokazuje najbolje performanse prilikom predikcije trajanja sonanata.

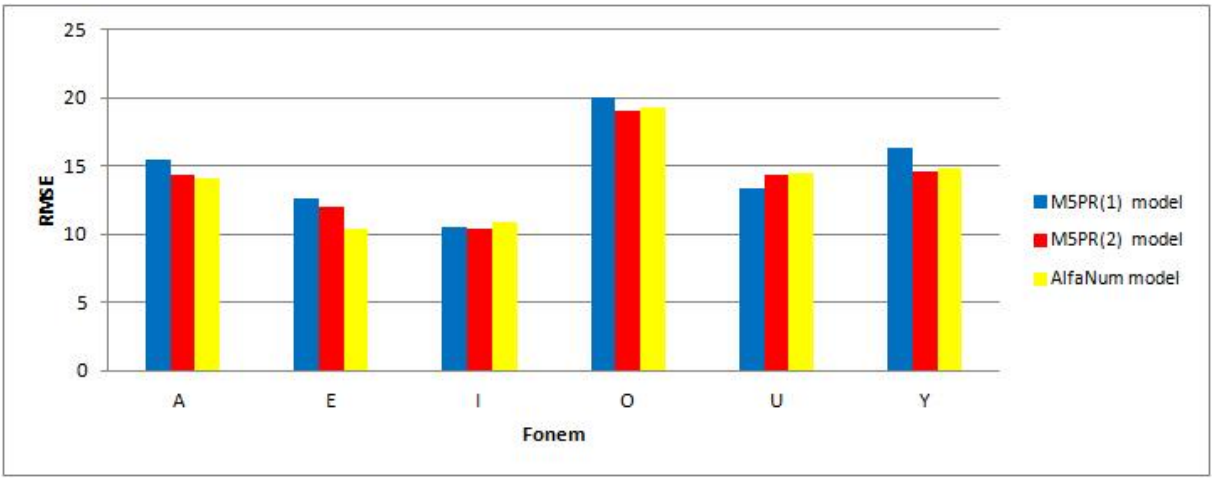
Fonem	**M5PR(1) model**	**M5PR(2) model**	**AlfaNum model**
Co	13,707	13,338	19,929
Ce	14,214	12,311	8,887
CCo	6,890	8,394	15,174
CCe	13,701	12,525	13,448
CHo	15,086	8,695	8,022
CHe	11,607	17,665	11,913
Afrikati	12,784	12,643	13,077

Tabela 7.21 – Greška predikcije trajanja afrikata kod različitih modela

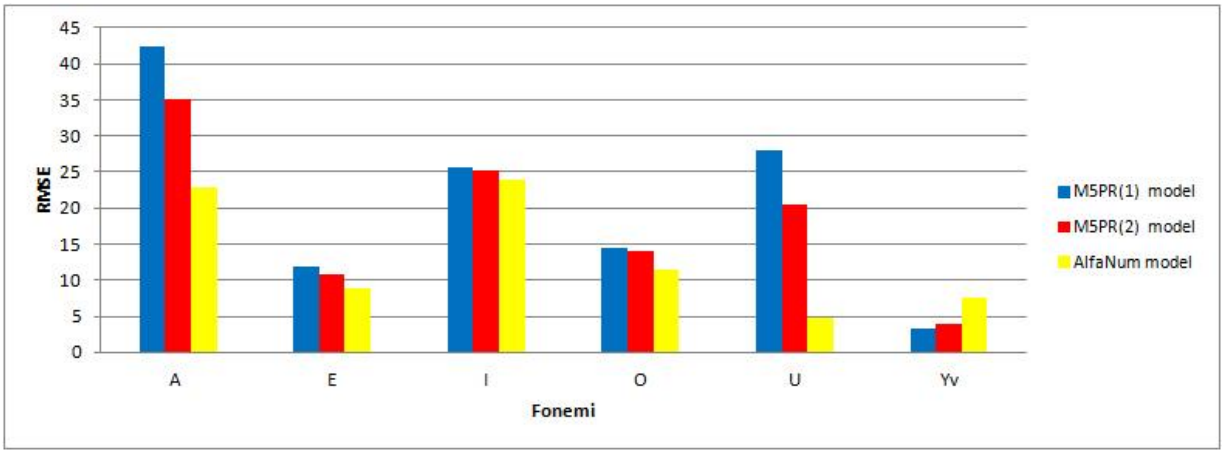
Fonem	**M5PR(1) model**	**M5PR(2) model**	**AlfaNum model**
M	14,777	15,580	20,313
N	15,160	16,941	15,738
NJ	26,525	6,178	3,079
L	9,107	11,182	10,091
R	9,043	9,285	8,067
J	12,662	12,452	15,388
Sonanti	12,727	13,504	13,590

Tabela 7.22 – Greška predikcije trajanja sonanata kod različitih modela

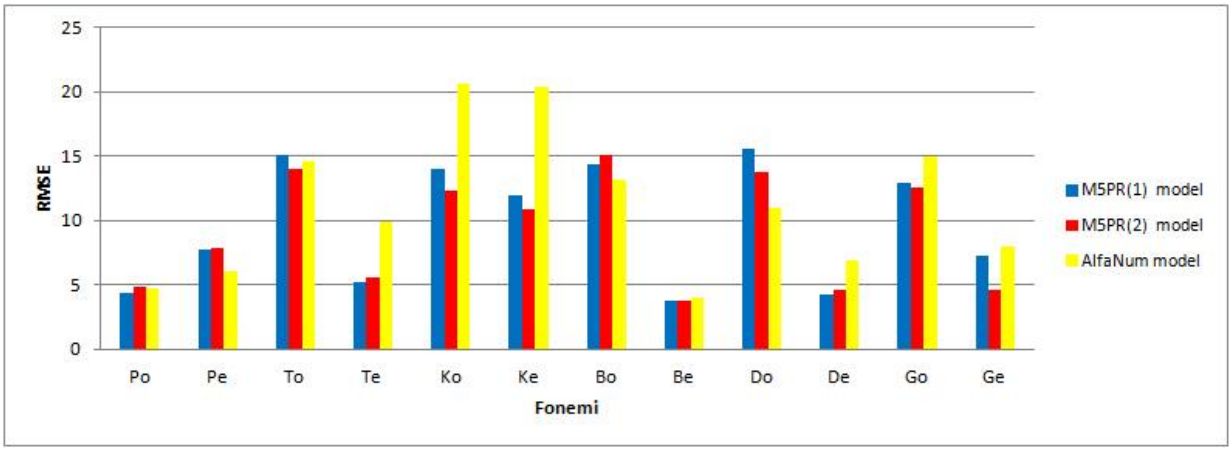
Na slikama 7.39 - 7.44 dat je uporedni prikaz vrednosti greške predikcije, odnosno RMSE koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela trajanja za različite grupe fonema. U okviru određene grupe fonema histogramom su prikazane vrednosti greške predikcije za svaki od fonema koji pripada datoj grupi a pojavljuje se u okviru test rečenica koje nisu učestvovale u fazi obuke modela.



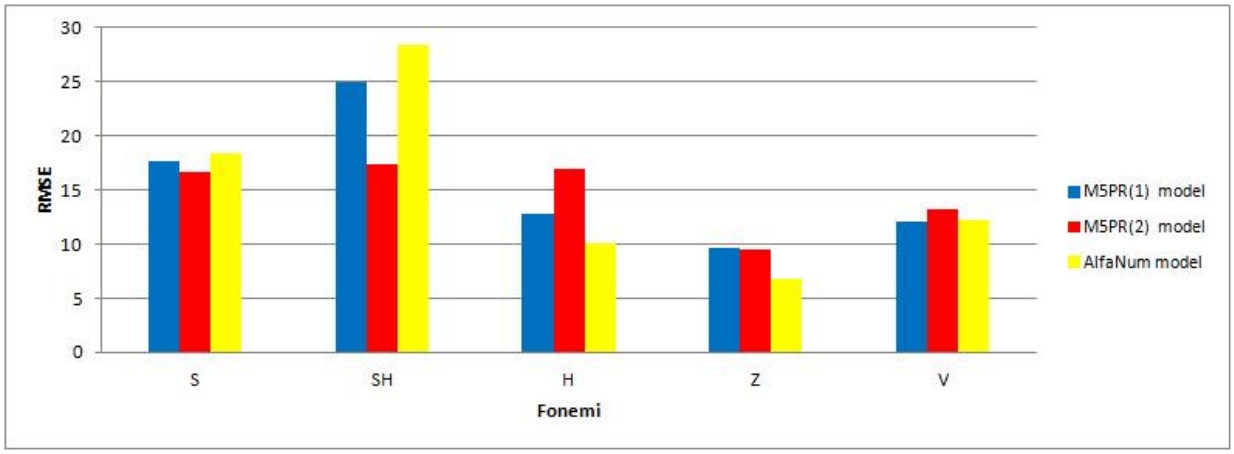
Slika 7.39 – Uporedni prikaz greške predikcije trajanja nenaglašenih vokala kod različitih modela trajanja



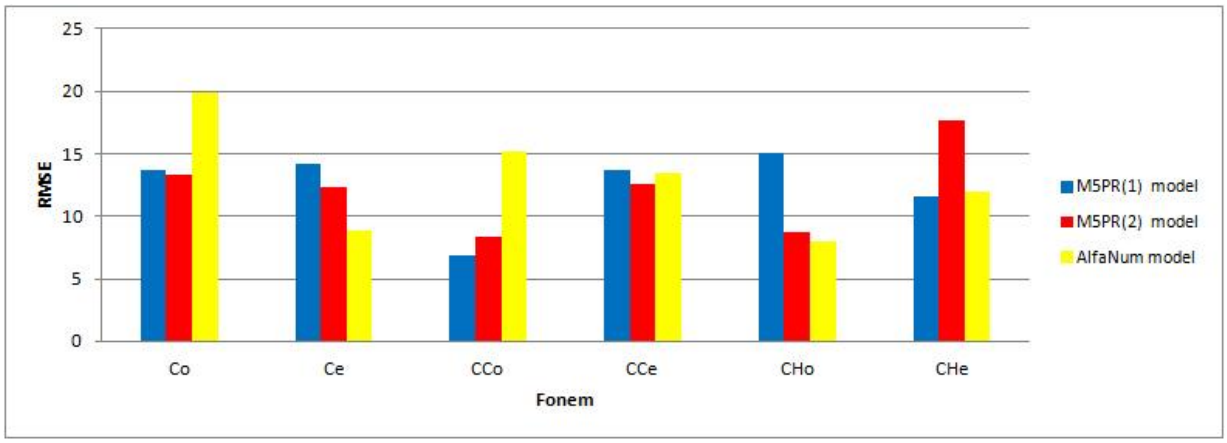
Slika 7.40 – Uporedni prikaz greške predikcije trajanja naglašenih vokala kod različitih modela trajanja



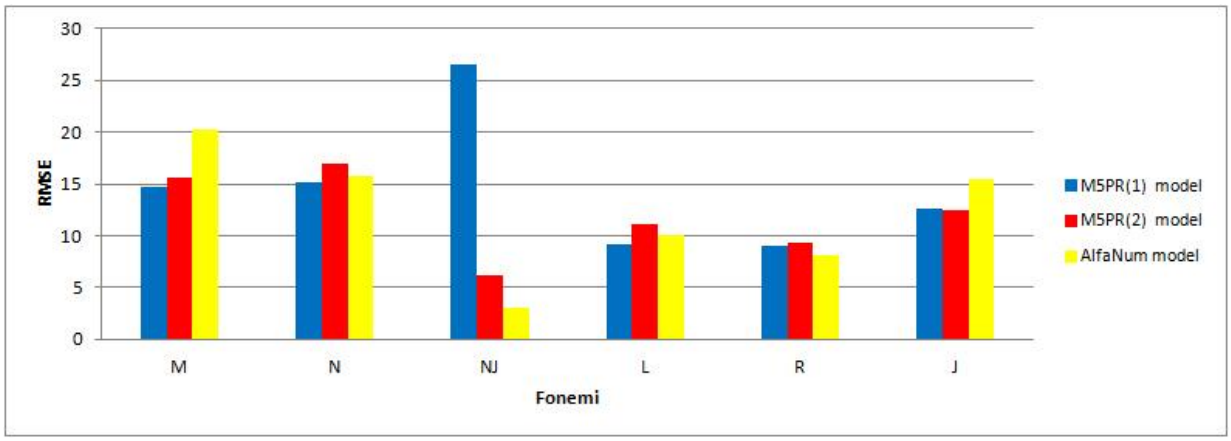
Slika 7.41 – Uporedni prikaz greške predikcije trajanja ploziva kod različitih modela trajanja



Slika 7.42 – Uporedni prikaz greške predikcije trajanja frikativa kod različitih modela trajanja

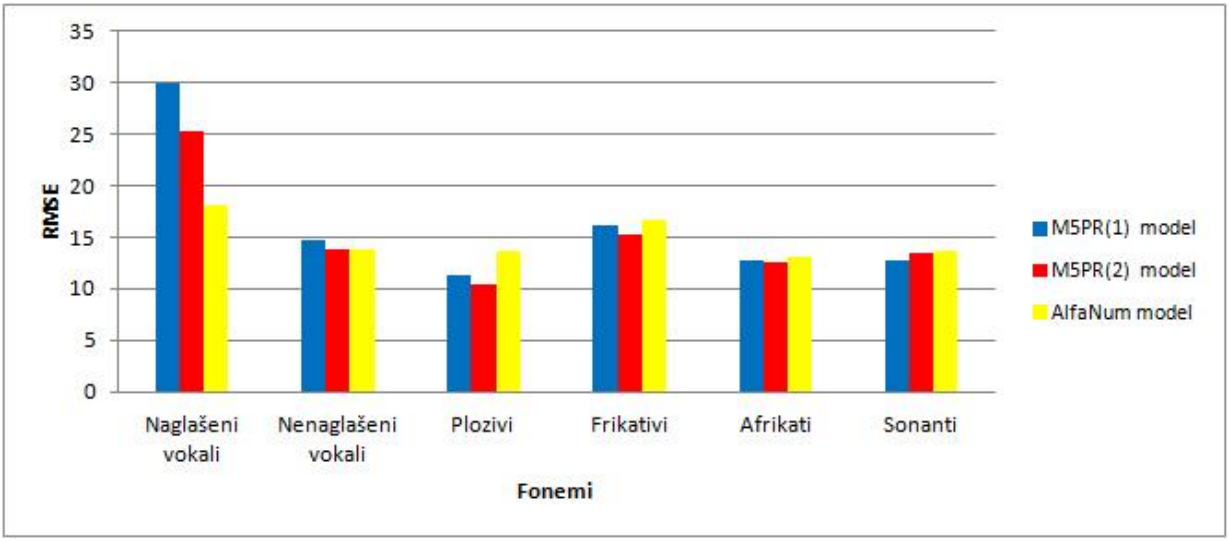


Slika 7.43 – Uporedni prikaz greške predikcije trajanja afrikata kod različitih modela trajanja



Slika 7.44 – Uporedni prikaz greške predikcije trajanja sonanata kod različitih modela trajanja

Na slici 7.45 dat je uporedni prikaz vrednosti RMSE koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela trajanja za svaku grupu fonema.



Slika 7.45 – Uporedni prikaz greške predikcije trajanja različitih grupa fonema kod različitih modela trajanja

Rezultati testiranja na rečenicama koje nisu učestvovale u fazi obučavanja modela pokazuju da su prediktivne performanse M5PR (2) modela bolje nego M5PR (1) modela, odnosno da je greška predikcije trajanja naglašenih i nenaglašenih vokala manja ukoliko se primeni model trajanja vokala prilikom predikcije trajanja. Takođe, greška predikcije je manja i ako se u slučaju predikcije trajanja ploziva, frikativa i afrikata primeni model trajanja konsonanata umesto modela trajanja glasova.

8. ZAKLJUČAK

S obzirom na značaj trajanja govornih segmenata sa perceptivnog stanovišta, odnosno na ulogu trajanja u razumevanju izgovorenog teksta, specijalizovani modul za određivanje potrebnog trajanja predstavlja komponentu TTS sistema od izuzetne važnosti za proizvodnju sintetizovanog govora visokog kvaliteta. Modelovanje trajanja govornih segmenata u različitim jezicima jeste predmet mnogobrojnih do sada sprovedenih istraživanja u kojima su primenjivane različite tehnike modelovanja. Razvoj matematičkog modela trajanja glasova u automatskoj sintezi govora na srpskom jeziku primenom metoda automatskog učenja na osnovu obimnog korpusa snimljenog govora, kao i identifikovanje najuticajnijih faktora na trajanje glasova u srpskom jeziku predstavljaju predmet istraživanja u okviru ove doktorske disertacije.

Prilikom modelovanja trajanja glasova u srpskom jeziku korišćena je obimna govorna baza koja sadrži približno 2000 rečenica i 16000 reči. U procesu modelovanja trajanja neophodna komponenta TTS sistema, koja prethodi modulu za određivanje trajanja određenog govornog segmenta u datom kontekstu, jeste modul za automatsko generisanje odgovarajućeg vektora obeležja kojim se predstavlja svaki fonem u govornoj bazi. Elementi vektora obeležja opisuju određeni govorni segment i kontekst u kome se on nalazi, pri čemu je vrednost svakog obeležja zapravo jedan od mogućih nivoa faktora koji utiče na trajanje govornog segmenta.

Faktori koji su korišćeni u procesu modelovanja trajanja glasova u srpskom jeziku u okviru ove disertacije su identitet, vrsta, način artikulacije (za konsonante) i mesto artikulacije (za konsonante) trenutnog segmenta; vrsta, način artikulacije (za konsonante) i zvučnost prethodnog i narednog segmenta; položaj trenutnog segmenta u slogu; naglašenost sloga kome pripada dati segment i vrsta akcenta; položaj sloga u reči; vrsta reči u kojoj se nalazi trenutni segment; dužina reči; fokus reči i položaj reči u frazi.

Prilikom razvoja modela trajanja glasova u srpskom jeziku u okviru ove disertacije korišćeni su različiti algoritmi softverskog paketa WEKA (Hall et al., 2009) među koje se ubrajaju LinearRegression (LR), M5P, M5PR, REPTree, kao i meta algoritmi AditiveRegression (AR), Bagging (BAG) i Stacking (STACK). Pomenuti modeli razvijani su primenom odgovarajućeg algoritma na obimnom govornom korpusu koji sadrži 98214 glasova, odnosno 38543 vokala i 59671 konsonant.

Modeli trajanja razvijeni su za celokupan skup fonema u srpskom jeziku. Takođe, razvijeni su i posebni modeli za vokale, odnosno konsonante. Evaluacija razvijenih modela

trajanja glasova realizovana je pomoću postupka ukrštene validacije, kao i na podacima koji nisu bili korišćeni u fazi obuke modela. U ovom slučaju celokupna govorna baza bila je podeljena na dva dela. Skup koji je korišćen za obuku modela sadrži 80% govorne baze dok je testiranje modela vršeno na preostalih 20%. Poređenje razvijenih modela izvršeno je na osnovu kvantitativnih pokazatelja kao što su RMSE, MAE i CC između stvarne i predviđene vrednosti trajanja glasova.

U okviru ove disertacije prikazane su vrednosti kvantitativnih pokazatelja RMSE, MAE i CC na osnovu kojih je izvršeno poređenje razvijenih modela trajanja glasova u koje se ubrajaju LR, M5P, M5PR, REPTree sa i bez potkresivanja stabla, AR M5PR, AR REPTree, BAG M5PR, BAG REPTree i STACK M5PR. Na osnovu dobijenih rezultata može se zaključiti da najbolje prediktivne performanse poseduje AR REPTree model, dok je LR model najlošiji model za predikciju trajanja glasova u srpskom jeziku. Takođe je utvrđeno da je M5PR algoritam bolji kao osnovni nego REPTree kod Bagging algoritma za razliku od aditivne regresije gde je REPTree evidentno bolji kao osnovni algoritam.

Kao i kod modela trajanja glasova najbolje performanse ima model trajanja konsonanata razvijen primenom aditivne regresije, ali u ovom slučaju bolji rezultati su dobijeni ukoliko se kao osnovni algoritam koristi M5PR. AR REPTree model trajanja vokala poseduje najbolje prediktivne performanse za predikciju trajanja vokala u srpskom jeziku. Primena linearne regresije prilikom razvoja modela trajanja kako konsonanata, tako i vokala ponovo daje najlošije rezultate.

U cilju poboljšanja ostvarenih performansi razvijenih modela izvršeno je uklanjanje iz govorne baze onih glasova koji u najvećoj meri doprinose grešci predikcije (engl. *outliers*). Na osnovu dobijenih rezultata može se zaključiti da AR REPTree model trajanja glasova predstavlja model najboljih prediktivnih performansi, kao i da primena AR REPTree modela trajanja konsonanata, odnosno vokala takođe daje najbolje rezultate prilikom predikcije trajanja konsonanata, odnosno vokala. Nakon uklanjanja glasova, konsonanata, odnosno vokala u blizini granica opsega kod AR REPTree modela ostvareno je smanjenje RMSE od 5,13% kod modela trajanja glasova, 4,69% kod modela trajanja konsonanata, odnosno 4,21% kod modela trajanja vokala.

Istraživanje je takođe obuhvatilo poređenje prediktivnih performansi M5PR modela razvijenog u okviru ove disertacije za srpski jezik i modela zasnovanih na stablima odluke koji su razvijeni za predikciju trajanja glasova u turskom (Öztürk, 2005), češkom (Batušek, 2002), korejskom (Lee & Oh, 1999), srpskohrvatskom (Sečujski et al., 2011), hindu i telugu (Krishna & Murthy, 2004) jeziku, kao i poređenje sa rezultatima koji su dobijeni primenom regresionih

stabala za razvoj modela trajanja konsonanata, odnosno vokala u litvanskom (Norkevičius & Raškinis, 2008), grčkom (Lazaridis et al., 2011) i engleskom (Lazaridis et al., 2011) jeziku. Na osnovu dobijenih rezultata može se uočiti da su prediktivne performanse M5PR modela trajanja uporedljive ili čak prevazilaze performanse modela koji su razvijeni za druge jezike a bili su dostupni autoru.

Prilikom poređenja prediktivnih performansi LR, M5P, M5PR, AR REPTree i BAG M5PR modela trajanja konsonanata, odnosno vokala utvrđeno je da modeli razvijeni u okviru ove disertacije za srpski jezik poseduju bolje performanse nego modeli razvijeni za grčki, odnosno engleski jezik (Lazaridis et al., 2011).

Evaluacija razvijenih modela trajanja, zasnovanih na regresionim stablima, kao i poređenje dobijenih rezultata sa rezultatima AlfaNum (Sečujski et al., 2011) modela trajanja glasova izvršena je testiranjem modela na dve grupe test rečenica. U prvoj grupi nalaze se rečenice koje su učestvovale prilikom razvoja modela u fazi obučavanja modela, dok drugu grupu čine rečenice koje nisu učestvovale u obučavanju modela. U okviru ove disertacije izvršeno je poređenje vrednosti RMSE koje su dobijene primenom M5PR (1), M5PR (2) i AlfaNum modela. Model M5PR (1) predstavlja model trajanja glasova, dok M5PR (2) podrazumeva primenu modela trajanja vokala, odnosno konsonanata u zavisnosti od toga koji je fonem u pitanju. Kod AlfaNum modela primenjen je poseban model trajanja za svaki od fonema. Na osnovu dobijenih rezultata nakon testiranja modela na obe grupe test rečenica može se zaključiti da su prediktivne performanse M5PR (2) modela bolje nego M5PR (1) modela, odnosno da je greška predikcije trajanja naglašenih i nenaglašenih vokala manja ukoliko se primeni model trajanja vokala prilikom predikcije trajanja. Takođe, greška predikcije je manja i ako se u slučaju predikcije trajanja ploziva, frikativa ili afrikata primeni model trajanja konsonanata umesto modela trajanja glasova.

Na osnovu rezultata testiranja modela na rečenicama koje nisu učestvovale u fazi obuke modela može se zaključiti da su realne prediktivne performanse modela trajanja vokala razvijenog u okviru ove disertacije i AlfaNum modela približno iste prilikom predikcije trajanja nenaglašenih vokala, dok AlfaNum model pokazuje bolje performanse u predikciji trajanja naglašenih vokala. Takođe, može se zaključiti da model trajanja konsonanata razvijen u okviru ove disertacije poseduje najbolje prediktivne performanse u predikciji trajanja ploziva, frikativa i afrikata. Prilikom predikcije trajanja sonanata greška predikcije trajanja je najmanja ukoliko se primeni M5PR (1) model trajanja glasova.

U okviru ove doktorske disertacije razvijeno je više modela trajanja glasova, odnosno vokala i konsonanata u srpskom jeziku primenom različitih metoda automatskog učenja a takođe

je izvršeno i njihovo međusobno poređenje na osnovu odgovarajućih kvantitativnih pokazatelja. Nakon implementacije u postojeći sintetizator govora (Sečujski et al., 2007) dalje istraživanje moglo bi da obuhvati subjektivnu evaluaciju razvijenih modela u cilju ocene kvaliteta sintetizovanog govora na osnovu kvalitativnih pokazatelja kao što su razumljivost i prirodnost govora. Takođe, jedno od vrlo interesantnih područja za dalje istraživanje jeste i detaljnija analiza modela trajanja vokala i konsonanata, koja bi obuhvatila analizu uticaja određenih faktora na pojedine grupe glasova i utvrđivanje optimalnog skupa atributa, a potom i modela trajanja za svaku grupu glasova.

Rezultati istraživanja dobijeni u ovoj doktorskoj disertaciji predstavljaju samo jedan korak ka konačnom cilju, poboljšanju kvaliteta sintetizovanog govora, ali svakako veoma značajan, imajući u vidu zavisnost govornih tehnologija od jezika s jedne strane, kao i perspektivnost, značaj i mogućnosti primene sinteze govora s druge strane.

LITERATURA

[1] J. Allen, M.S. Hunnicutt, D. H. Klatt, *From Text to Speech: the MITalk System*, Cambridge University Press, Cambridge, UK, 1987.

[2] J. Bakran, *Zvučna slika hrvatskoga govora*, IBIS grafika, Zagreb, Hrvatska, 1996.

[3] K. Bartkova, C. Sorin, "A Model of Segmental Duration for Speech Synthesis in French", *Speech Communication*, vol. 6, pp. 245-260, 1987.

[4] R. Batušek, "A Duration Model for Czech Text-to-Speech Synthesis", *in Proc. of Speech Prosody 2002*, France, pp. 167-170, 2002.

[5] P. Boula de Mareüil, "Phonetic-prosodic Transcription of Italian for Text-to-Speech Synthesis", in *Proc. of VI International Congress SILFI 1999*, Duisburg, Germany, 1999.

[6] C. Bouzon, D. Hirst, "The Influence of Prosodic Factors on the Duration of Words in British English", in *Proc. of Speech Prosody*, Aix-en-Provence, France, pp. 191-194, 2002.

[7] L. Breiman, J. H. Fredman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees*, Wadsworth Statistics/Probability Series, Belmont, CA., 1984.

[8] L. Breiman, "Bagging Predictors", *Journal of Machine Learning*, vol. 24, pp. 123-140, 1996.

[9] I. Bulyko, M. Ostendorf, P. Price, "On the Relative Importance of Different Prosodic Factors for Improving Speech Synthesis", in *Proc. of ICPhS*, vol. 1, San Francisco, pp. 81-84, 1999.

[10] W. N. Campbell, *Multi-level Speech Timing Control*, PhD dissertation, University of Sussex, 1992.

[11] R. Carlson, B. Granstrom, "A Search for Durational Rules in Real Speech Database", *Phonetica*, vol. 43, pp. 140-154, 1986.

[12] M. Chen, "Vowel Lenght Variation as a Function of the Voicing of the Consonant Environment", *Phonetica*, vol. 22, 1970.

[13] H. C. Choi, R. W. King, "On the Use of Spectral Transformation Speaker Adaptation in HMM Based Isolated-Word Speech Recognition", *Speech Communication*, vol. 17, , pp. 131-143, 1995.

[14] H. Chung, "Duration Models and the Perceptual Evaluation of Spoken Korean", in *Proc. of Speech Prosody*, Aix-en-Provence, France, pp. 219-222, 2002.

[15] H. Chung, M. Huckvale, "Linguistic Factors Affecting Timing in Korean with Application to Speech Synthesis", in *Proc. of EUROSPEECH*, Denmark, 2001.

[16] A. Cohen, I. H. Slis, J. Hart, "On Tolerance and Intolerance in Vowel Perception", *Phonetica*, vol. 16, 1967.

[17] T. Crystal, A. House, "Segmental durations in connected speech signals", *Journal of the Acoustical Society of America*, vol. 83, issue 4, pp. 1553-1573, 1988.

[18] V. Delić, M. Sečujski, N. Jakovljević, M. Janev, R. Obradović, D. Pekar, "Speech Technologies for Serbian and Kindred South Slavic languages", *Advances in Speech Recognition*, InTech, pp. 141-164, 2010.

[19] V. Delić, M. Sečujski, N. Jakovljević, M. Gnjatović, I. Stanković, "Challenges of Natural Language Communication with Machines", Chapter 19 in *DAAAM Int. Scientific Book 2013*, Eds: B. Katalinic & Z. Tekic, Vienna, Austria, pp. 371-388, 2013.

[20] V. Delić, M. Sečujski, M. Bojanić, D. Knežević, N. Vujnović Sedlar, R. Mak, "Aids for the Disabled Based on Speech Technologies - Case Study for the Serbian Language", XI Int. Conf. ETAI-2013, Ohrid, Macedonia, pp. E2-1.1-4, 2013.

[21] G. Epitropakis, D. Tambakas, N. Fakotakis, G. Kokkinakis, "Duration modelling for the Greek Language", in *Proc. of EUROSPEECH 1993*, Berlin, Germany, pp. 1995-1998, 1993.

[22] A. Febrer, J. Padrell, A. Bonafonte, "Modeling Phone Duration: Application to Catalan TTS", in *Proc. of 3rd ESCA/COCOSDA Workshop on Speech Synthesis*, Australia, pp. 43-46, 1998.

[23] J. L. Flanagan, *Speech analysis, synthesis and perception*, Springer-Verlag, Berlin, 1972.

[24] O. Goubanova, S. King, "Bayesian Networks for Phone Duration Prediction", *Speech Communication*, vol. 50, no. 4, pp. 301-311, 2008.

[25] J. Gros, *Samodejno tvorjenje govora iz besedil*, Založba ZRC, Ljubljana, Slovenia, 2000.

[26] S. Gudurić, D. Petrović, "O prirodi glasa [r] u srpskom jeziku", *Zbornik Matice srpske za filologiju i lingvistiku*, vol. 48, no. 1-2, pp. 135-150, 2005.

[27] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA data mining software: An update", *SIGKDD Explorations*, vol. 11, issue 1, 2009.

[28] N. Iwahashi, Y. Sagisaka, "Statistical Modeling of Speech Segment Duration by Constrained Tree Regression", *IEICE Trans. Inform. Systems*, E83-D(7), pp. 1550-1559, 2000.

[29] S. Jovičić, M. Popović, M. Đorđević, "Akustičke karakteristike vokala u šapatu", ETRAN, Budva, 1996.

[30] S. Jovičić, *Govorna komunikacija, fiziologija, psihoakustika i percepcija*, Nauka, Beograd, Srbija, 1999.

[31] M. Kaariainen, T. Malinen, "Selective Rademacher Penalization and Reduced Error Pruning of Decision Trees", *Journal of Machine Learning Research*, vol. 5, pp. 1107-1126, 2004.

[32] E. Klabbers, *Segmental and Prosodic Improvements to Speech Generation*, PhD dissertation, Technical University of Eindhoven, Netherlands, 2000.

[33] N. Kaiki, K. Takeda, Y. Sgisaka, "Statistical Analysis for Segmental Duration Rules in Japanese Speech Synthesis", in *Proc. of ICSLP'90*, pp. 17-20, 1990.

[34] D. H. Klatt, "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence", *Journal of the Acoustical Society of America*, vol. 59, pp. 1209-1221, 1976.

[35] D. Klatt, "Review of Text-to-Speech Conversion for English", *Journal of the Acoustical Society of America*, vol. 82, no. 3, pp. 737- 783, 1987.

[36] K. J. Kohler, "Zeitstrukturierung in der Sprachsynthese", *ITG-Tagung Digitale Sprachverarbeitung*, vol. 6, pp. 165-170, 1988.

[37] Đ. Kostić, *Metodika izgradnje govora u dece oštećena sluha*, Savez društva defektologa Jugoslavije, Beograd, 1971.

[38] N. S. Krishna, H. A. Murthy, "Duration Modeling of Indian Languages Hindi and Telugu", in *5-th ISCA Speech Synthesis Workshop*, Pittsburgh, USA, pp. 197-202. 2004.

[39] N. S. Krishna, P. P. Talukdar, K. Bali, A. G. Ramakrishnan, "Duration Modeling for Hindi Text-to Speech Synthesis System", in *Proc. of ICSLP*, pp. 789–792, 2004.

[40] A. Lazaridis, P. Zervas, N. Fakotakis, G. Kokkinakis, "A CART Approach for Duration Modeling of Greek Phonemes", in *Proc. of SPECOM*, Moscow, Russia, pp. 287-292, 2007.

[41] A. Lazaridis, P. Zervas, G. Kokkinakis, "Segmental Duration Modeling for Greek Speech Synthesis", in *Proc. of IEEE ICTAI-2007*, Patras, Greece, pp. 518-521, 2007.

[42] A. Lazaridis, V. Bourna, N. Fakotakis, "Comparative Evaluation of Phone Duration Models for Greek Emotional Speech", *Journal of Computer Science*, vol. 6, no. 3, pp. 341-349, 2010.

[43] A. Lazaridis, I. Mporas, T. Ganchev, G. Kokkinakis, "Improving Phone Duration Modeling Using Support Vector Regression Fusion", *Speech Communication*, vol. 53, no.1, 2011.

[44] S. Lee, Y-H. Oh, "Tree-based Modeling of Prosoding Phrasing and Segmental Duration for Korean TTS system", *Speech Communication*, vol. 28, issue 4, pp. 283-300, 1999.

[45] I. Lehiste, P. Ivić, *Prozodija reči i rečenice u srpskohrvatskom jeziku*, Izdavačka knjižarnica Zorana Stojanovića, Sremski Karlovci, 1996.

[46] A. M. Liberman, M. Studdert-Kennedy, "Phonetic Perception", in R. Held, H. Leibowitz, H. L. Tehner (eds.), *Handbook of Sensory Physiology*, Springer Verlag, 1979.

[47] B. Lindblom, "Spectrographic Study of Vowel Reduction", *Journal of the Acoustical Society of America*, vol. 35, pp. 1773-1781, 1963.

[48] A. Malecot, "Acoustic Cues for Nasal Consonants. An Experimental Study Involving a Tape-splicing Technique", *Language*, vol. 32, pp. 274-284, 1956.

[49] M. Marković, T. Milićev, "The effect of rhythm unit length on the duration of vowels in Serbian", in *Proc. of 19th ISTAL (International Symposium of Theoretical and Applied Linguistics*), Thessaloniki, Greece, pp. 305-313, 2009.

[50] M. Marković, I. Bjelaković, "Kvalitet pretoničnih vokala u govoru Novog Sada", *Godišnjak Filozofskog fakulteta u Novom Sadu*, XXXVI-1, pp. 99-111, 2011.

[51] M. Mihaljević, *Generativna i leksička fonologija*, Školska knjiga, Zagreb, 1991.

[52] H. Mixdorff, D. T. Nguyen, N. T. Wu, "Duration Modeling in a Vietnamese Text-to-Speech System", in *Proc. of SPECOM 2005*, Patras, Greece, 2005.

[53] B. Moebius, J. P. H. van Santen, "Modeling Segmental Duration in German Text-to-Speech Synthesis", in *Proc. of International Conference on Spoken Language Processing*, Philadelphia, USA, vol. 4, pp. 2395-2398, 1996.

[54] G. Norkevičius, G. Raškinis, "Modeling Phone Duration of Lithuanian by Classification and Regression Trees, Using Very Large Speech Corpus", *Informatica,* vol. 19, no. 2, pp. 271-284, 2008.

[55] D. K. Oller, "The Effect of Position in Utterance on Speech Segment Duration in English", *Journal of the Acoustical Society of America*, vol. 54, pp. 1235-1247, 1973.

[56] Ö. Öztürk, *Modeling phoneme durations and fundamental frequency contours in Turkish speech,* PhD dissertation, Middle East Technical University, 2005.

[57] G. E. Peterson, H. L. Barney, "Control Method Used in a Study of the Vowels", *Journal of the Acoustical Society of America*, vol. 24, pp. 175-184, 1952.

[58] D. Petrović, S. Gudurić, *Fonologija srpskoga jezika*, Institut za srpski jezik SANU, Beogradska knjiga, Matica srpska, Beograd, 2010.

[59] K. S. Rao, B. Yegnanarayana, "Modeling Syllable Duration in Indian Languages using Support Vector Machines", in *Proc. of ICISIP 2005*, India, pp. 258-263, 2005.

[60] K. S. Rao, B. Yegnanarayana, "Modeling Durations of Syllables Using Neural Networks", *Computer Speech and Language*, vol. 21, no. 2, pp. 282-295, 2007.

[61] M. Riley, "Tree-based Modeling of Segmental Durations", *Talking Machines: Theories, Models and Designs*, Elsevier, 1992.

[62] J. P. H. van Santen, "Contextual Effects on Vowel Duration", *Speech Communication*, vol. 11, no. 6, pp. 513-546, 1992.

[63] J. P. H. van Santen, "Timing in Text-to-Speech Systems", *in Proc. of EUROSPEECH,* Berlin, Germany, pp. 1397-1404, 1993.

[64] J. P. H. van Santen, "Exploring N-way Tables with Sums-of-Product Models", *Journal of Mathematical Psychology*, vol. 37, no. 3, pp. 327-371, 1993.

[65] J. P. H. van Santen, "Assignment of Segmental Duration in Text-to-Speech Synthesis", *Computer Speech and Language*, vol. 8, pp. 95-128, 1994.

[66] J. P. H. van Santen, "Computation of Timing in Text-to-Speech Synthesis", in W. B. Kleijn and K. K. Paliwal (editors), *Speech Coding and Synthesis*, Elsevier, 1995.

[67] J. F. Schouten, R. J. Ritsma, B. Lopez Cardozo, "Pitch of the Residue", *Journal of the Acoustical Society of America*, vol. 34, pp. 1418-1424, 1962.

[68] M. Sečujski, *Prozodijski elementi u sintezi govora na srpskom jeziku*, Magistarki rad, Fakultet tehničkih nauka, Novi Sad, 2002.

[69] M. Sečujski, V. Delić, D. Pekar, R. Obradović, D. Knežević, "An Overview of the AlfaNum Text-to-Speech Synthesis System", in *Proc. of SPECOM*, Moscow, Russia, pp. 3-7, 2007.

[70] M. Sečujski, N. Jakovljević, D. Pekar, "Automatic Prosody Generation for Serbo-Croatian Speech Synthesis Based on Regression Trees", in *Proc. of INTERSPEECH 2011*, Florence, Italy, pp. 3157-3160, 2011.

[71] A. R. M. Simoes, "Predicting Sound Segment Duration in Connected Speech: An Acoustical Study of Brazilian Portuguese", in *Proc. of Workshop on Speech Synthesis*, Autrans, France, pp. 173-176, 1990.

[72] D. J. Sharf, T. Hemeyer, "Identification of Place of Consonant Articulation from vowel Formant Transitions", *Journal of Acoustical Society of America*, vol. 51, pp. 652-658, 1972.

[73] W. Slawson, "Vowel Quality and Musical Timbre as Function of Spectrum Envelope and Fundamental Frequency", *Journal of the Acoustical Society of America*, vol. 43, pp. 87-101, 1968.

[74] S. Sovilj-Nikić, *Trajanje vokala kao jedan od prozodijskih elemenata u sintezi govora na srpskom jeziku*, Magistarski rad, Fakultet tehničkih nauka, Novi Sad, 2007.

[75] Ž. Stanojčić, Lj. Popović, S. Mićić, *Gramatika srpskog jezika*, Zavod za udžbenike i nastavna sredstva, Beograd, Srbija, 2005.

[76] C. J. Stone, "Additive Regression and other Nonparametric Models", *Annals of Statistics*, vol. 13, pp. 689-705, 1985.

[77] M. Šipka, *Kultura govora*, Prometej, Novi Sad, 2008.

[78] M. Telebak, *Pravogovor*, Prometej, Novi Sad, 2011.

[79] M. Vainio, *Artificial Neural Network Based Prosody Models for Finnish Text-to-Speech Synthesis*, Academic Dissertation, University of Helsinki, Faculty of Arts, Department of Phonetics, Finland, 2001.

[80] J. J. Venditti, J. P. H. van Santen, "Modeling Vowel Duration for Japanese Text-to-Speech Synthesis", in *Proc. of International Conference on Spoken Language Processing*, Sydney, Australia, 1998.

[81] S. Vladisavljević, *Poremećaji izgovora*, Privredni pregled, Beograd, 1981.

[82] Z. Weibin, S. Liqin, N. Xiaochuan, “Duration Modeling for Chinese Synthesis from C-ToBI Labeled Corpus”, in *Proc. of ICSLP 2000*, vol. 3, Beijing, China, pp. 159-162, 2000.

[83] L. White, *English Speech Timing: A Domain and Locus Approach*, PhD dissertation, University of Edinburgh, 2002.

[84] H. I. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kauffman Publishing, San Francisco, 2005.

[85] J. Yamagishi, H. Kawai, T. Kobayashi, “Phone Duration Modeling Using Gradient Tree Boosting”, *Speech Communication*, vol. 50, no. 5, pp. 405-415, 2008.

[86] M-S. Yu, N-H. Pan, M-J. Wu, “A Statistical Model with Hierarchical Structure for Predicting Prosody in Mandarin Text-to-Speech System”, *Journal of Chinese Institute of Engineers*, vol. 28, no. 3, pp. 385-399, 2005.

BIOGRAFIJA

Sandra Sovilj-Nikić rođena je 23.05.1978. godine u Novom Sadu. Diplomirala je na Fakultetu tehničkih nauka – elektrotehnički odsek, smer telekomunikacije 23.10.2003. godine sa prosečnom ocenom 9,54 (od 10,00) odbranivši diplomski rad pod nazivom "Primena genetskog algoritma u median filtriranju". Po završetku osnovnih studija upisala je poslediplomske studije na Katedri za telekomunikacije i obradu signala na Fakultetu tehničkih nauka u Novom Sadu. Sve ispite predviđene planom i programom poslediplomskih studija položila je sa prosečnom ocenom 10,00. Bila je stipendista Novosadskog univerziteta i Rotary kluba tokom osnovnih studija, a 2001. godine dobila je i jednokratnu stipendiju Vlade Kraljevine Norveške. Tokom poslediplomskih studija bila je stipendista Ministarstva za nauku i zaštitu životne sredine Republike Srbije. Kao stipendista Ministarstva sarađivala je na projektima "Razvoj komunikacione infrastrukture za funkcionalne mreže zasnovan na digitalnoj obradi signala i komunikacionom softveru", "Razvoj govornih tehnologija za srpski jezik i primena u Telekomu" i "Govorna komunikacija čovek – mašina". Dana 10.07.2007. godine odbranila je svoju magistarsku tezu pod nazivom "Trajanje vokala kao jedan od prozodijskih elemenata u sintezi govora na srpskom jeziku". Polje njenog interesovanja su govorne tehnologije, naročito sinteza govora na osnovu teksta, što je ujedno i disciplina kojoj pripadaju njena magistarska teza i doktorska disertacija, a takođe i oblast veštačke inteligencije i primena različitih metoda automatskog učenja u rešavanju optimizacionih problema i problema numeričke predikcije. Govori engleski i italijanski jezik, služi se nemačkim jezikom.