Sveučilište u Zagrebu

Fakultet elektrotehnike i računarstva

Gordan Kreković

Specifičnosti sustava za pretvorbu teksta u govor za hrvatski jezik

Zagreb, 2010.

Ovaj rad izrađen je na Zavodu za elektroničke sustave i obradu informacija pod vodstvom prof. dr. sc. Davora Petrinovića i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2009./2010.

Sadržaj

1. Uvod

Govor, kao prirodan, jednostavan i brz način komuniciranja, ponekad je najprikladniji oblik izmjene informacija između čovjeka i računala. Da bi računalo bilo sposobno formirati govornu poruku i uputiti je čovjeku, ono mora imati ugrađen sustav za umjetnu tvorbu govora koji iz pisanog ulaza može generirati valni oblik govornog signala. Umjetna se tvorba govora često naziva i sintezom govora. Na tehnologiji sinteze govora temelje se alati za pomoć osobama s različitim vrstama tjelesnih nedostataka, sustavi za pružanje informacija putem telefona ili pametnih kioska, uređaji na koje korisnik tijekom rada ne smije ili ne može skretati pogled, programska podrška za učenje stranih jezika, sustavi za dijalog čovjeka i računala, mnogi proizvodi zabavne industrije i drugo. Cjeloviti sustavi za pretvorbu teksta u govor (eng. *text-to-speech system*) mogu raditi sa slobodnim tekstovima u kojima se pojavljuju kratice, brojevi, adrese elektroničke pošte i razne druge oznake. Sve elemente teksta prvo treba pretvoriti u riječi kako bi se iz njih generirali strukturirani fonemski nizovi s pripadnim oznakama za način izgovora. Iz tako dobevnog simboličkog zapisa zatim se generira valni oblik govornog signala primjenom neke od postojećih metoda sinteze. Svaka od metoda ima svoje posebnosti u smislu razumljivosti i prirodnosti izgovora, fleksibilnosti, složenosti i zahtjeva vezanih uz sklopovsko i programsko okruženje. Na odabir metode u konkretnoj primjeni utječe cilj koji se želi ostvariti sustavom za sintezu. Danas se najviše koriste artikulacijska, formantna i statistička parametarska sinteza te sinteza ulančavanjem (eng. *concatenative synthesis*). Sinteza ulančavanjem zajednički je naziv za sve metode kod kojih se sintetički govor tvori povezivanjem prethodno snimljenih segmenata prirodnog govora. Kako na karakter pojedinog glasa utječu njemu susjedni glasovi, izbor segmenata duljine jednog fonema nije prikladan. Zato su se istraživanja lančane sinteze usmjerila na jedinice poput slogova, trifona i difona. Koncept difonske sinteze polazi od pojednostavljene i ne uvijek sasvim istinite pretpostavke da koartikulacijski efekti ne prelaze više od dva fonema. Ipak, difonska je sinteza zbog jednostavnosti baze snimljenih segmenata, vrlo stabilne tranzicije između fonema i razmjerno visoke razumljivosti sintetiziranog govora vrlo popularna i često korištena metoda.

Postići prirodnost sintetizranog govora najveći je problem sinteze ulančavanjem. Naime, simljeni segmenti mogu biti preuzeti iz različitih okruženja i imati različite značajke poput intonacije i glasnoće. Povezivanjem takvih isječaka nastaju izobličenja zbog kojih rezultantni sintetizirani govor djeluje neprirodno. Imati na raspolaganju isječke, koji dobro odgovaraju jedan uz drugoga za svaki izbor fonemskog niza, riješilo bi problem, ali bi baza segmenata tada bila vrlo kompliciranа za

izradu, zauzimala bi mnogo memorijskog prostora i zahtijevala bi složenije algoritme odabira segmenata prilikom njihovog ulančavanja. Modifikacija prozodije segmenata iz tih je razloga prikladnije rješenje. Neke su od tehnika za spajanje difonskih jedinica i modifikaciju prozodije metoda periodom sinkronoziranog preklapanja i zbrajanja signala (eng. *pitch synchronous overlap and add method*) i linearna predikcija pobuđena ostatkom (eng. *residual excited LPC*). Upravo te tehnike koristi difonski sintetizator višejezičnog sustava *Festival Speech Synthesis System* [1]. Osim potpunog sintetizatora govora, *Festival* omogućuje ugradnju novih govorničkih glasova i novih jezika, a zbog brojnih programskih sučelja, pogodno je okruženje za istraživanje različitih metoda sitneze.

Ovaj rad opisuje prvi cjeloviti sustav za pretvorbu teksta u govor za hrvatski jezik razvijen u okruženju *Festival*. Izrada sustava obuhvatila je pripremu, snimanje i označavanje baze difona te razvoj potrebnih jezičnih modula. Poznat nam je još samo jedan pokušaj realizacije sintetizatora govora za hrvatski jezik temeljen na metodi sinkroniziranog preklapanja i zbrajanja signala [2]. Međutim, taj se rad samo bavi izgradnjom difonske baze, dok naš sustav dodatno donosi izvorna rješenja vezana uz normalizaciju teksta i leksičku analizu. Također, za prikupljanje i označavanje difona, koristili smo sasvim različite metode.

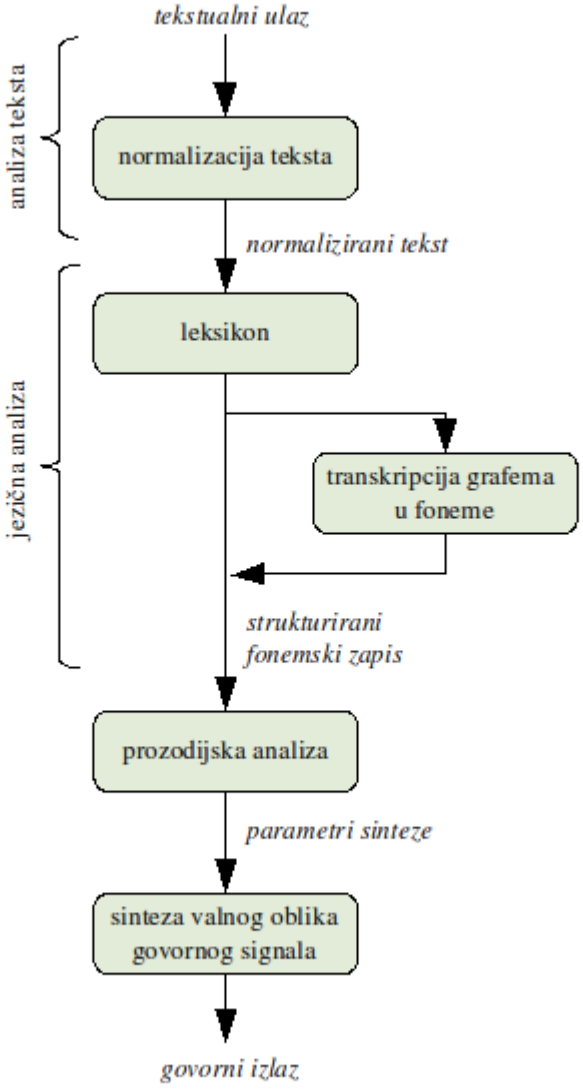
Slična istraživanja vezana uz difonsku sintezu provedena su za poljski [3], slovenski [4], korejski [5], škotski gaelski [6], njemački [7] i druge jezike.

Ispitivanja kvalitete izrađenih sustava predstavljaju još jedan izazov područja sinteze govora. Dosljedno i relevantno ispitivanje teško je provesti zbog nedostatka općenitog dogovora oko njegove strukture i sadržaja. Ipak, ponavljanjem postupaka drugih istraživača može se uspostaviti kriterij za usporedbu različitih sustava. U sklopu ovoga rada provedeno je mjerenje kvalitete postojećim pristupima.

2. Sustavi za sintezu govora

2.1. Struktura sustava za pretvorbu teksta u govor

Sinteza govora općenito se odnosi na postupak pretvorbe pisanog ulaza u govorni izlaz. Ulaz može biti u obliku grafemskog, ortografskog ili fonemskog zapisa, ali i neprilagođeni slobodni tekst. Sustavi za pretvorbu teksta u govor uključuju postupke normalizacije teksta, jezične analize, prozodijske analize i generiranja valnog oblika govornog signala.



Slika 1: *Generalizirala struktura sustava za pretvorbu teksta u govor*

U ulaznom tekstu mogu se pojaviti jedinice za koje se ne može načiniti fonemska transkripcija izravnom primjenom pravila. To su brojevi, datumi, sati, kratice, adrese elektroničke pošte i webskih sjedišta te općenito znakovni nizovi koji sadrže neslovne simbole. Prvi korak u procesu sinteze govora odnosi se na raspisivanje takvih jedinica u riječi. Postupak se naziva normalizacijom ili predobradbom teksta, a vrši ga modul za analizu teksta koji iz proizvoljnog tekstualnog ulaza stvara listu riječi. Pritom se najčešće koristi skup rukom zadanih pravila. Normalizaciju je važno provesti prije ostalih koraka zato što sve raspisane riječi podliježu pravilima izgovora i moraju biti obrađene jezičnim modulom. Također, na taj se način izbjegavaju moguće višesmislenosti poput, primjerice, tretiranja kratica riječima koje prati točka za kraj rečenice. Postupak normalizacije ne pokriva sve slučajeve jer je broj različitih oblika znakovnih nizova, koji se mogu pojaviti, vrlo velik, a dodatne probleme stvaraju neočekivane pogreške u tekstu i nejednoznačnosti koje niti čovjek ne bi znao ispravno interpretirati.

Zadatak modula za jezičnu analizu pretvorba je normaliziranog teksta u odgovarajući zapis izgovora koji obuhvaća fonemski zapis, slogovnu strukturu i oznaku naglaska. Postupak se vrši pretraživanjem leksikona u kojem se uz svaku riječ nalazi eksplicitni zapis izgovora. Ako tražena riječ nije sadržana u leksikonu, na njoj se mogu primijeniti skupovi pravila koji opisuju zamjene grafema fonemima, glasovne promjene, formiranje slogovne strukture i određivanje naglasaka.

Cilj modula za prozodijsku analizu stvoriti je parametre za sintezu koji će dati najprirodniji izgovor zadanog sadržaja. Pod pojmom prozodija podrazumijevaju se fonetski učinci poput glasnoće, visine glasa, stanki u govoru, brzine i ritma.

2.2. Primjene

Tehnologija umjetne tvorbe govora već se dugo vremena koristi za pomoć osobama s različitim motoričkim ili osjetilnim poteškoćama. Čitači zaslona oblik su namjenske programske podrške čiji je zadatak interpretirati sadržaj zaslona. Slijepim, slabovidnim i nepismenim osobama takvi alati omogućuju rad na osobnom računalu i pristup pisanim sadržajima u elektroničkom obliku. Formiranje, sintezu i reprodukciju željene poruke ostvaruju posebni prijenosni uređaji koji pomažu nijemim osobama pri govornoj komunikaciji. Takvi su uređaji redovito opremljeni namjenskim tipkovnicama koje služe za vrlo brzo slaganje rečenica. U posljednje vrijeme, tehnologije sinteze govora sve više primjenjuju osobe s disleksijom i drugim poteškoćama pri čitanju.

Govorna poruka katkad je efikasnija ili prikladnija od pisane. Zbog toga se sintetizatori govora

ugrađuju u mjerne ili kontrolne sustave, a redovito se koriste u situacijama kada ruke i oči korisnika trebaju ostati slobodne kao što je slučaj s navigacijskim sustavima za cestovna vozila.

Osobine sustava za pretvorbu teksta u govor čine ga izvrsnim laboratorijskim alatom za znanstvena istraživanja u lingvistici. Potpuna kontrola parametara i ponovljivost pokusa pogodna su obilježja za ispitivanje intonacijskih modela koji će doprinijeti još većoj kvaliteti sintetiziranog govora.
U zabavnoj industriji potreba za sintezom govora pojavljuje se pri izradi animiranih filmova i računalnih igara, a jednostavni se sintetizatori sve češće ugrađuju i u dječje igračke. Na tržištu također postoje programski paketi za učenje i usavršavanje stranih jezika sposobni generirati pravilan izgovor riječi i fraza.
Sprega tehnologija za prepoznavanje i sintezu govora omogućuje dvosmjernu izmjenu informacija između čovjeka i računala koja je zbog intuitivnosti i brzine katkad poželjnija od one ostvarene tradicionalnim korisničkim sučeljima. Kombiniraju li se govorne tehnologije sa sustavima umjetne inteligencije i bazama podataka, mogu se ostvariti sustavi za dijalog čovjeka i računala. Krajnji je cilj takvih sustava u potpunosti zamijeniti ljudskog operatora u raznovrsnim uslugama pružanja informacija. Tako se sve više automatizira odgovaranje na upite putem telefona koji su najčešće vezani uz redove voženje, telefonske brojeve, vremenske prognoze i razne informacije pozivnih centara. Jedna je zanimljiva primjena pametni kiosk (eng. *smart kiosk*) postavljen na javnom mjestu sa svrhom pružanja informacija turistima. Tako bi turist u stranom gradu mogao razgovorom s računalom doznati potrebne podatke, primjerice, o znamenitostima, smještaju, restoranima i ostalim sadržajima.

3. Okruženje za razvoj sintetizatora govora

Festival Speech Synthesis System višejezični je sustav za sintezu govora razvijen u Centru za istraživanje govorne tehnologije na Sveučilištu u Edinburghu [1]. Osim mogućnosti cjelovite pretvorbe teksta u govor, *Festival* nudi brojna programska sučelja pa tako služi kao okolina za ugradnju novih jezika i glasova govornika, a također je pogodan za istraživanje različitih tehnika sinteze govora i ugradnju novih modula. Sustav je napisan u programskom jeziku C++ i sadrži interpreter naredbi baziran na skriptnom jeziku *Scheme* koji je vrlo važan za razvojnu okolinu. *Scheme* je neproceduralni funkcionalni jezik pogodan za obradbu znakovnih nizova, odlučivanje na temelju skupa pravila i ostale primjene koje se pojavljuju prilikom razvoja modula za nove jezike. Za *Festival* je dosad razvijen veći broj jezika poput engleskog, španjolskog i poljskog. Uz postojeću definiciju jezika, prilično je jednostavno ugraditi novi govornički glas. Primjerice, želimo li sintetizator za engleski jezik upogoniti novim glasom, to možemo lako učiniti prateći postupak opisan u dokumentaciji sustava. Najčešće samo treba snimiti i obraditi bazu difona te prilagoditi parametre sintetizatora valnog oblika, intonacijskog modela i modela trajanja fonema karakteristikama snimljenog glasa. Mnogo složenija situacija razvoj je cjelovitog sustava koji obuhvaća definiciju novog jezika. Izazovi takvog pristupa nisu samo pisanje specifičnih normalizacijskih i jezičnih modula, nego i oblikovanje baze difona, označavanje segmenata u snimljenom materijalu i mnogi drugi.

3.1. Razvoj sintetizatora govora za novi jezik

Prilikom ugradnje novog jezika i govorničkog glasa u okruženje *Festival*, potrebno je obaviti sljedeći proces [8]:

1. Definirati korišteni skup fonema
2. Oblikovati bazu difona
3. Snimiti i označiti sadržaj baze
4. Izgraditi modul za analizu teksta
5. Napisati leksikon i pravila za tvorbu izgovora
6. Definirati trajanja fonema
7. Izgraditi intonacijski model
8. Ispitati kvalitetu izrađenog sustava

U sklopu izrade sintetizatora za hrvatski jezik provedeni su svi opisani postupci osim izgradnje

intonacijskog modela.

3.1.1. Definicija skupa fonema

Jezični moduli, razne pomoćne skripte, koje se koriste prilikom oblikovanja baze i automatskog označavanja snimljenog materijala, kao i modul za sintezu valnog oblika koriste definiciju skupa fonema za svoj rad. Zato je za svaki novi jezik ili narječje prvo potrebno izraditi popis fonema i njihov opis. Opis je ostvaren nabrajanjem obilježja kao što su vrsta fonema (suglasnik ili samoglasnik), mjesto tvorbe suglasnika i druge klasifikacije glasova. Obilježja su također definirana unutar istog modula što osigurava konzistentnost i kompatibilnost s ostalim dijelovima sustava.

Festival podržava postojanje više skupa fonema istovremeno i omogućuje preslikavanje među skupovima. Svaka definicija skupa fonema sadrži naziv skupa, listu obilježja i listu fonema.

```
  (defPhoneSet
     NAZIV
     LISTA_OBILJEŽJA
     LISTA_FONEMA )
```

Obilježa su određena skupom vrijednosti koje mogu poprimiti. Tako je, primjerice, obilježje zvučnosti suglasnika određeno podjelom na zvučne, bezvučne i zvonke glasovi.

```
(
     (vc + -)                                  ;; samoglasnik ili suglasnik
     (zvucnost zv bezv zvonki 0)               ;; obilježje zvučnosti
      ...
)
```

U listi fonema uz svaki simbol stoje vrijednosti obilježja redoslijedom definicije obilježja. Prema prethodnom primjeru, uz simbol *b* stajat će znak minus, koji govori da se radi o suglasniku, oznaka *zv*, koja klasificira fonem u skupinu zvučnih, te redom ostatak vrijednosti obilježja.

Lista fonema sadržava i oznake tišine (eng. *silence phones*). Difoni, koji u sebi sadrže pauzu, zapravo su fonemi na samom početku ili samom kraju riječi.

3.1.2. Oblikovanje difonske baze

Cilj izrade difonske baze sakupiti je valne oblike govornog signala svih potrebnih difona. Oni se izrezuju iz snimaka prirodnog govora. Sadržaj snimljenog materijala treba pomno pripremiti kako bi pokrivao sve željene difone.

Difoni se mogu prikupiti iz stvarnih riječi ili iz strateški izgrađenih besmislenih riječi, takozvanih logatoma. Kada na raspolaganju stoji dovoljno velika količina kvalitetno snimljenog kontinuiranog govora, ona se može koristiti u svrhe izrade baze. No, često je potrebno imati barem nekoliko sati snimljenog govora da bi pokrivenost difona bila zadovoljavajuća. Također, pojave poput nestalnosti izgovora, naglasaka, izražene intonacije i varijacije intenziteta loše utječu na konačnu kvalitetu sintetiziranog govora. Zato se, kada za to postoji mogućnost, u kontroliranim i stalnim uvjetima snima izgovor planiranog skupa riječi.

Logatomi su nizovi glasova koji ne moraju nužno imati značenje, a sastavljeni su kako bi ubrzali i unaprijedili proces izgradnje difonske baze. Lista difona generira se kombiniranjem fonema u parove, a zatim se svaki difon ugrađuje u kontekst kojim se olakšava izgovor te omogućuje puna i jasna artikulacija. Da bi se smanjili nepoželjni prijelazni efekti na početku i kraju riječi, ciljni se difon smiješta u sredinu logatoma. Prednosti korištenja ovog pristupa su brojne: otklonjena je potreba za traženjem stvarnih riječi koje sadrže sve željene difone, govornik je manje podložan pogrešnom izgovoru nego pri čitanju stvarnih riječi, a označavanje granica fonema mnogo je lakše bilo da se ono provodi ručno ili automatski.

3.1.3. Snimanje i označavanje materijala za difonsku bazu

Osobine govornika važni su čimbenici koji utječu na kvalitetu difonske sinteze. Izbor govornika nije lak zadatak i često obuhvaća razmatranje mnogo različitih aspekata. Osoba bi trebala biti izvorni govornik bez poteškoća u govoru i čitanju, sposobna jasno i pravilno artikulirati pripremljene riječi. Bez posebne pripreme može biti vrlo teško dugo govoriti stalnim intenztitetom. Zato je poželjno snimanje obaviti s profesionalnim govornikom naviknutim na takve situacije. Konzistentnost izgovora od velike je važnosti jer se prilikom lančane sinteze različiti dijelovi govora povezuju. Osnovno znanje iz fonetike i govornih tehnologija čini govornika svjesnim čemu treba pokloniti više pažnje, kakav izgovor je poželjan i zašto je tome tako. Treba izbjegavati snimanje u situacijama kada je govornik prehlađen ili promukao iz drugih razloga.

Općenite preporuke za kvalitetno snimanje su sljedeće:

- tiho okruženje za snimanje poput tonskog studija ili namjenske sobe za snimanje
- stalni uvjeti snimanja: nepromjenjiva udaljenost mikrofona od govornika, iste postavke uređaja za snimanje, minimalne varijacije u intezitetu izgovora

- izgovor bez naglasaka i varijacije u tempu
- snimanje cjelovitog sadržaja u jednom terminu jer je iste uvjete teško ponoviti kasnije; ako nije moguće sve snimiti odjednom, preporuča se snimanje u istim razdobljima dana

Iz snimljenog materijala potom treba izdvojiti difone. To se obavlja označavanjem granica fonema. Programski paket *Festvox* sadrži mnoge pomoćne alate za izradu difonske baze iz snimljenog sadržaja pa tako nudi i mogućnost automatskog označavanja difona u logatomima [9]. Za automatsko označavanje koristi se algoritam dinamičke vremenske deformacije (eng. *dynamic time warping*) koji poravnava snimke s postojećim prethodno označenim izgovorima i na odgovarajući način preslikava oznake granica fonema. Takva metoda ne daje uvijek zadovoljavajuće rezultate, pogotovo kada se provodi za logatome koji nemaju otprije označeni izgovor. Automatski postavljene oznake mogu se ručno popraviti što je daleko jednostavnije nego ispočetka ručno označavati čitav snimljeni materijal. Prilikom ručnog označavanja, valja se oslanjati na iskustvo i pomoć namjenskih alata poput *Emulabela*. Više o problematici automatskog i ručnog označavanja usmjerenoj na izgradnju hrvatskog sintetizatora, piše u sljedećem poglavlju.

3.1.4. Analiza teksta

Normalizacija ulaznog teksta prvi je stupanj jezične obradbe pri kojemu se ulazne rečenice prevode u odgovarajuće liste riječi. Tijekom tog postupka donose se odluke što učiniti sa znakovima interpunkcije i ostalim neslovnim tekstualnim simbolima. Identificiraju se elementi poput kratica, brojeva, datuma, adresa elektroničke pošte te se raspisuju u potpune ortografske znakovne nizove. Nastoji se pokriti što veći broj očekivanih oblika.

Primjer normalizacije proveden na rečenici:

> *Maksimalna temperatura zraka od 27.2 °C zabilježena je 9.8.2004. godine u 13:43.*

trebao bi rezultirati nizom:

> *maksimalna temperatura zraka od dvadeset sedam cijelih dva stupnja celzija zabilježena je devetog kolovoza dvije tisuće četvrte godine u trinaest sati i četrdeset tri minute*

Za pravilnu pretvorbu nije uvijek dovoljan samo izdvojeni tekstualni element, već i njegov kontekst. U mnogim jezicima brojevi su promjenjive riječi pa ih treba postaviti u odgovarajući padež, rod i broj. Također, kada je kratica jednoslovna ili predstavlja homograf postojećoj riječi, to treba prepoznati u sklopu normalizacije teksta. Složenost ovog podsustava ovisi o jeziku, ali i razini

detaljnosti ugrađenih pravila. Analiza teksta često se prilagođava namjeni sintetizatora. Ako je očekivano da će sintetizator najviše čitati, primjerice, novinske članke, normalizacijski modul bit će izgrađen tako da pokriva što više oblika koji se inače pojavljuju upravo u takvom tipu teksta.

Kratice, koje ne tvore neku drugu riječ i imaju nedvojbeno značenje, ne raspisuju se u procesu normalizacije, već je njihov cjeloviti izgovor sadržan u leksikonu. Primjeri su *str.* (stranica), *itd.* (i tako dalje) i *prof.* (profesor).

U okruženju *Festival* normalizacijski modul za novi jezik treba razviti otpočetka. Identifikaciju oblika tekstualnih jedinica olakšava mogućnost korištenja regularnih izraza.

3.1.5. Leksikon i pravila za tvorbu izgovora

Za svaku riječ generiranu jezičnim modulom potrebno je imati izgovor. Izgovor je definiran odgovarajućim fonemskim nizom struktiriranim u slogove i oznakama naglaska na svakom pojedinom slogu. Leksikon je podstustav koji za traženu riječ iz definiranog skupa pruža eksplicitni izgovor. Pojedini zapis, uz riječ i njezin izgovor sadrži i jednostavnu oznaku koja dodatno opisuje riječ pa može služiti, primjerice, za razlikovanje homografe. Takva se oznaka definira po volji, a može označavati vrstu riječi ili njezinu ulogu u rečenici.

Struktura zapisa je lista čiji je prvi član riječ predstavljena znakovnim nizom, drugi član nedjeljiva pomoćna oznaka, a treći cjeloviti izgovor. Tipični zapisi su, primjerice:

```
(lex.add.entry '("sinus" nn (((s i) 1) ((n u s) 0))))
(lex.add.entry '("stribor" nn (((s t r i) 1) ((b o r) 0))))
```

Rad ovog podsustava podrazumijeva pretraživanje leksikona i pružanje izgovora za traženu riječ. Ako riječ nije sadržana, za nju se primjenjuju pravila za generiranje izgovora. Prilikom planiranja razvoja jezičnih modula valja razmisliti koje riječi treba staviti u leksikon. Često je rješenje leksikonom osigurati točan izgovor što većeg broja riječi, a pravila izgovora koristiti samo kao dodatni podsustav koji će se pobrinuti za riječi izuzete iz leksikona. Takvo rješenje je optimalno za one jezike koji imaju izrazito složen skup pravila izgovora, kao što je engleski jezik. S druge strane, kada se pravila mogu eksplicitno ispisati, prikladno je imati manji leksikon koji sadrži samo jednoslovčane simbole, kratice, homografe, riječi stranog izgovora i slične iznimke, a tvorbu izgovora prepustiti podsustavu za provođenja pravila.

Osnovni podsustav za provođenje pravila izgovora u *Festivalu*, iako jednostavan, omogućuje izgradnju razmjerno složenih skupova pravila. Općeniti je oblik pravila [9]:

```
(LS [ alpha ] DS => beta)
```

Kada je ulazni znakovni niz `alpha` okružen s lijeve strane sadržajem `LS` i s desne strane sadržajem `DS`, na izlaz se bilježi znakovni niz `beta`. Niz `alpha` mora sadržavati barem jedan znak, dok `LS`, `DS` i `beta` mogu biti prazni. Primjerice, sljedeće pravilo opisuje pretvorbu samoglasnika /r/ u slogotvorni suglasnik /r̩/:

```
(pau [ r ] SG => rr)
```

Oznaka *pau* predstavlja granicu riječi, dok je SG skup svih suglasnika. Dakle, prema ovom pravilu, kada se fonem /r/ nađe na početku riječi, a slijedi ga suglasnik, on postaje /r̩/.

Za svaki ulazni simbol redom se provjerava svako pravilo iz liste. Ako uvjeti odgovaraju, pravilo se izvršava, zapisuje se izlaz te se ponovno pretražuje lista pravila od početka.

4. Pretvorba teksta u govor za hrvatski jezik

Ovo poglavlje opisuje izvorna rješenja ugrađena u naš sustav za pretvorbu teksta u govor. Sadržani su svi elementi sustava pri čemu su posebno naglašene specifičnosti zbog hrvatskog jezika.

4.1. Definicija skupa fonema

Proučavanjem fonetičke literature [10] i razmatranjem praktičnih aspekata, odlučili smo koristiti sljedeći skup fonema: /a/, /b/, /c/, /č/, /ć/, /d/, /ʒ/, /ʒ́/, /e/, /f/, /g/, /h/, /i/, /i̯/ , /k/, /l/, /l̩/, /m/, /n/, /ń/, /o/, /p/, /r/, /r̩/, /s/, /š/, /t/, /u/, /v/, /z/ i /ž/. Iz cjelovitog skupa hrvatskih fonema isključen je samoglasnik /ie/ koji odgovara realizaciji difona *i-e*.

Prilikom izbora fonema, razmatrano je proširenje osnovnog skupa fonemima iz stranih jezika. Kako se u tekstu mogu očekivati riječi, za čiji ispravan izgovor nije dovoljan osnovni skup fonema niti ugrađena pravila izgovora, ignoriranje stranih fonema u suvremenim sustavima za sintezu govora nije dopustivo. S druge strane, proširenjem skupa fonema, brzo raste broj difona što otežava postupke oblikovanja, snimanja i označavanja baze. Također, govorniku nije uvijek jednostavno pravilno izgovoriti difone koji kombiniraju hrvatski i strani fonem. Zbog toga smo ipak odlučili zasada ne proširivati skup fonema, ali smo u jezični modul uključili pravila za preslikavanje stranih grafema u hrvatske foneme. Osim toga, eksplicitni izgovori najčešćih međunarodnih riječi dodani su u leksikon.

U konkretnoj realizaciji oznake fonema prilagođene su američkom standardnom znakovniku za razmjenu informacija (eng. *American Standard Code for Information Interchange, ASCII*) pa se tako, primjerice, za fonem /ʒ/ koristi višeslovna oznaka *dzx*, a za slogotvorni suglasnik /r̩/ oznaka *rr*. Oznaka tišine je *pau*.

4.2. Oblikovanje difonske baze

Prikupljanje difona pomoću generiranih logatoma nikad ranije nije provedeno za hrvatski jezik. Prilikom izgradnje našeg sustava, odlučili smo koristiti upravo tu metodu zbog brojnih prednosti opisanih u prethodnom poglavlju. Pogotovo je važna činjenica da je označavanje granica fonema na snimkama logatoma mnogo lakše nego na snimkama kontinuiranog govora za koje alati u sklopu *Festvox* paketa niti nemaju mogućnost automatskog označavanja.

Svaki se logatom sastoji od ciljnog difona i konteksta koji ga omeđuje, takozvanog difonskog nosača [11]. Svrha nosača osigurati je jasnu i punu artikulaciju te olakšati izgovor logatoma. Sadržajem i strukturom on mora biti posebno prilagođen situaciji pa se za različite razrede difona koriste različiti nosači. Zato je prije konstrukcije logatoma definirano šest razreda difona: samoglasnik-suglasnik (VC), suglasnik-samoglasnik (CV), suglasnik-suglasnik (CC), samoglasnik-samoglasnik (VV), slogotvorni suglasnik-suglasnik (SC) i suglasnik-slogotvorni slogotvorni (CS). Difoni su generirani kombiniranjem fonema iz odgovarajućih skupina. Primjerice, prilikom stvaranja razreda VC, svaki je samoglasnik uparen sa svim suglasnicima. Za pojedini razred potom je osmišljena odgovarajuća struktura logatoma.

Kada je to prikladno, u logatom se mogu uključiti po dva difona što smanjuje ukupni skup riječi za snimanje i obradbu. Takvo smo rješenje iskoristili za razrede CV i VC. Općeniti oblik logatoma koji sadrži CV i VC difone jest *ta*-c-v-c-*a* pri čemu *c* označava bilo koji samoglasnik, a *v* bilo koji suglasnik. Primjer pripadnika ove skupine riječ je *takuka* koja nosi difone *k-u* i *u-k*.

Strukture korištene za CC i VV razrede redom su *ta*-c1-c2-*ata* i *tat*-v1-v2-*ta*. Parovi dvaju jednakih fonema, isključeni su iz ovog skupa. Primjeri logatoma su *tarlata*, gdje je ciljni difon *r-l* i *tatueta* koji nosi par samoglasnika *u-e*.

Suglasnik /r/ postaje slogotvornim kada je okružen suglasnicima ili se nalazi na početku riječi, a slijedi ga suglasnik. Primjeri su riječi *rt*, *crta*, *krv* i *kvrga*. Za slučaj kada je slogotvorno /r̥/ na početku riječi, koristimo logatom *r-ta*. Njime smo pokrili tranziciju između tišine i slogtvornog /r̥/, odnosno difon *pau-r*. Kombinacije ostalih suglasnika sa fonemom /r̥/ ostvaruju se u logatomima oblika *tat-r*-c-*a* and *ta*-c-*r-ta*. Primjeri pripadnika ovih skupina su *tatrva* i *tavrta*. Razmatrano je stavljanje SC i CS difona u isti logatom, ali je na kraju ideja odbačena jer su neke kombinacije teško izgovorljive.

Slučajevi fonema na granici riječi također moraju biti pokriveni. Za suglasnike na početku, odnosno na kraju riječi korištene su redom strukture c-*ata* i *tata*-c, a za samoglasnike v-*ta* i *tat*-v.

Opisani algoritam za generiranje logatoma ostvaren je u skriptnom jeziku *Scheme*. Na početku su definirane konteksti za pojedine razrede difona koje koriste funkcije za konstrukciju logatoma. Funkcije jednostavno iteriraju foneme iz odgovarajućih skupina i fonemski par postavljaju u pripremljeni kontekst. U nastavku je kao primjer prikazana funkcija za generiranje logatoma oblika *ta*-c-v-c-*a*:

```
(define (list-cvcs)
  (apply
   append
   (mapcar
    (lambda (v)
      (mapcar
       (lambda (c)
       (list
        (list (string-append c "-" v) (string-append v "-" c))
        (append (car cvc-carrier) (list c v c) (car (cdr cvc-carrier)))))
       consonants))
    vowels)))
```

Korištenjem *Scheme* interpretera ugrađenog u okruženje *Festival*, izrađena je lista od 855 logatoma. Svaki redak liste sadržava oznaku logatoma, njegov fonetski zapis i ciljne difone.

```
...
( cro_0123 "pau t aa v u v aa pau" ("v-u" "u-v") )
( cro_0124 "pau t aa z u z aa pau" ("z-u" "u-z") )
( cro_0125 "pau t aa zx u zx aa pau" ("zx-u" "u-zx") )
( cro_0126 "pau t aa t a e t aa pau" ("a-e") )
( cro_0127 "pau t aa t a i t aa pau" ("a-i") )
...
```

4.3. Izrada difonske baze

4.3.1. Snimanje izgovora

Za snimanje izgovora odabrana je odrasla ženska osoba, izvorna govornica hrvatskog jezika sa znanjem fonetike i prethodnim iskustvom u javnom govorenju. Njezin zadatak bio je ispravno, čisto i ravnomjerno izgovoriti sve pripremljene logatome.

Snimanje je obavljeno u profesionalnom tonskom studiju tijekom jednog termina u trajanju od 3 sata. Materijali su pohranjeni u obliku digitalnog zvučnog zapisa 16-bitne rezolucije i frekvencije uzorkovanja 44.1 kHz. Za inicijalno snimanje izgovora, bilo je potrebno dva sata, dok je jedan sat utrošen na provjeru snimaka i ponavljanje nekih logatoma u svrhu povećanja kvalitete. Provjerom su utvrđene pogreške na oko 10% ukupnog broja inicijalno snimljenih logatoma. U većini slučajeva razlozi ponavljanja bili su pogrešan izgovor ili pretjerano naglašavanje.

Studijska snimka ručno je podijeljena tako da svaka datoteka zvučnog zapisa sadrži po jedan logatom. U tu svrhu je korišten alat za obradu zvuka *Audacity*. Svim zvučnim zapisima potom je

smanjena frekvencija uzorkovanja na 16 kHz zbog ograničenja korištene verzije sustava *Festival*.

Kako prilikom snimanja nije bio raspoloživ uređaj za elektrolaringografiju, oznake titranja glasnica postavljene su naknadno uz pomoć alata iz paketa *Festvox*.

4.3.2. Označavanje snimljenjog materijala

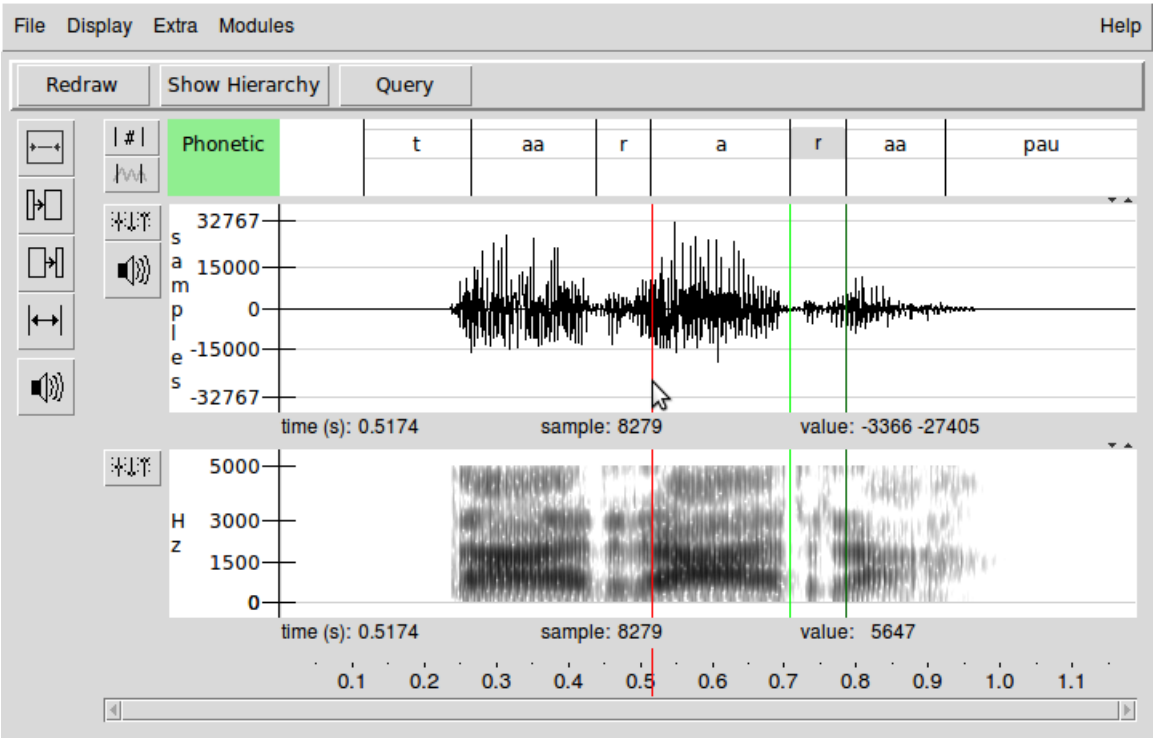
Prva faza označavanja snimljenog materijala automatsko je postavljanje oznaka za granice fonema. S obzirom da nije bilo sličnih pokušaja razvoja hrvatskog sintetizatora u okruženju *Festival* pa označeni izgovori pripremljenih logatoma za prikupljanje hrvatskih difona ne postoje, koristili smo sintezu za engleski jezik. U tu je svrhu trebalo napisati skriptu koja obavlja preslikavanje hrvatskih fonema u engleske. Njome je fonetski zapis naših logatoma prilagođen engleskom jeziku i omogućena je sinteza engleskim sintetizatorom. Korištenjem metode dinamičke vremenske deformacije, oznake granica preslikane su iz sintetiziranih izgovora na stvarne snimke. Zbog nepreciznosti postupka i velikih razlika među hrvatskim i engleskim fonemima, nismo očekivali dobre rezultate. Ipak se pokazalo da je automatsko označavanje vrlo djelotvorno čak i kod ovakvog međujezičnog postupka. Većina oznake je bila relativno uspješno postavljena te je osigurana zadovoljavajuća inicijalna razumljivost. Ipak, prilikom testiranja, u sintetiziranom govoru pojavljivali su se šumovi, izobličenja i oštre tranzicije koje su ostavljale dojam neugodnosti i povremeno činile dijelove govora nerazumljivim.

Ručna korekcija oznaka obavljena je uz pomoću alata *Emulabel* koji je dio sustava za manipulaciju i analizu govornih baza *EMU Speech Database System*. Za postizanje visoke kvalitete izlaznog govora, osobito je važno precizno postaviti granice među fonemima ciljnih difona. U tu svrhu pomaže prikaz valnog oblika u vremenskoj domeni, pripadni spektrogram i mogućnost reprodukcije dijelova govornog signala.

Kako se zbog fizikalnih ograničenja oblik ljudskog vokalnog trakta ne može trenutno promijeniti, niti prijelazi među glasovima ne mogu biti trenutni. U tom se smislu nije lako odlučiti za točan trenutak u kojem jedan glas završava, a drugi počinje. Prema iskustvu stečenom prilikom ručnog označavanja, granice treba tražiti na mjestima značajnijih promjena u spektrogramu i često pri nižim amplitudama signala. Najlošije automatski postavljene oznake bile su one na granicama samoglasnika i usnenih suglasnika.

Postupak ručnog popravljanja oznaka višestruko je dugotrajniji od svih ostalih koraka izrade difonske baze. Za efikasno dobivanje željenih rezultata potrebno je mnogo iskustva pa je zato

uputno pratiti perporuke drugih istraživača i razvojnika.



Slika 2: *Ručno označavanje granica fonema pomoću alata* Emulabel

4.4. Modul za analizu teksta

U trenutnoj fazi razvoja sustava, modul za analizu teksta može prepoznati i normalizirati brojeve, datume, oznake vremena, polusloženice i općenite znakovne nizove s neslovnim simbolima. U sklopu budućeg usavršavanja sustava svakako je predviđeno proširenje i poboljšanje funkcionalnosti. Kako je svrha sintetizatora čitati internetski sadržaj i interpretirati tekstualne informacije koje se pojavljuju prilikom uobičajenog rada na osobnom ili prijenosnom računalu, posebna je pažnja posvećena elementima poput adresa elektroničke pošte i webskih sjedišta, naziva datoteka s nastavcima i IP adresa.

4.4.1. Brojevi i oznake vremena

U hrvatskom standardnom jeziku, brojevi su promjenjive riječi. Količina se izriče glavnim brojevima, a mjesto u redoslijedu rednim. Od glavnih brojeva mijenjaju se *jedan*, *dva*, *tri* i *četiri* te složeni brojevi kojima je neki od navedenih posljednji član. Redni se brojevi mijenjaju kao određeni pridjevi. Rečenicu

Sastali su se predstavnici 34 zemalja po 2. put u 2010. godini.

pravilno bi bilo pročitati

Sastali su se predstavnici trideset četiriju zemalja po drugi put u dvije tisuće desetoj godini.

što znači da na temelju konteksta, modul za analizu teksta treba ispravno pogoditi rod, broj i padež. Programski ostvariti takav podsustav vrlo je teško.

Iz teksta se koriste dvije vrste informacija. Prvo, to je oblik tekstualnog elementa. Oblik sudjeluje u određivanju načina na koji se znamenkama i simbolima zapisan broj pravilno raspisuje u riječ. Neki od mogućih oblika su:

- niz dekadskih znamenaka od kojih je prva različita od nule – prirodni broj
- niz dekadskih znamenaka s predznakom – cijeli broj
- znakovni niz koji se sastoji od dekadskih znamenaka i točno jedne decimalne točke koja ne može biti zadnji znak u nizu; na prvom mjestu može biti predznak – decimalni broj
- prirodni broj koji završava točkom, a sljedeća riječ ne počinje velikim slovom – redni broj
- znakovni niz koji se sastoji od dekadskih znamenaka, znakova minus i kosih crta ili nizovi dekadskih znamenaka koji počinju nulom – telefonski broj
- posebni oblici: datumi, oznake vremena, poštanski brojevi, rezultati sportskih nadmetanja ili omjeri, IP adrese, itd.

Pravila, po kojima se identificiraju pojedini oblici, ne pokrivaju sve moguće situacije, ali su odabrana tako da u većini slučajeva daju ispravnu odluku.

Za pravilnu pretvorbu tekstualnog elementa, koji predstavlja ili sadrži broj, u niz riječi ponekad je dovoljno znati oblik. Tako se, primjerice, telefonski broj ne mijenja s obzirom na okolinu u kojoj se pojavljuje. Samo treba svaku znamenku posebno pretvoriti u jednostavni broj muškog roda u nominativu jednine. S druge strane, glavni i redni brojevi trebaju se po rodu, broju i padežu slagati s imenicom na koju se odnose. Zbog toga je potrebno znati koja od riječi u rečenici predstavlja tu imenicu i kojeg je ona roda, broja i padeža.

Vrsta i oblik riječi druga je informacija koju je potrebno izvući analizom teksta. U sustavu *Festival* problem se često rješava tako da se izgradi potpuni rječnik u kojem uz svaki zapis postoje svi relevantni podaci. Kako je objašnjeno u sljedećem poglavlju, gradnja rječnika osobito ne doprinosi

ukupnoj funkcionalnosti sustava za sintezu govora u drugim aspektima niti je obuhvaćena istraživačkim interesom. Dakako, eventualna proširenja sustava mogu uključiti rječnik s nekoliko tisuća ili nekoliko desetaka tisuća najučestalijih riječi u hrvatskom jeziku i tako osigurati pravilnu pretvorbu brojeva i ostalih tekstualnih elemenata, koji sadrže brojeve, u riječi pravilnog roda, broja i padeža. Za sada se određivanje oblika riječi obavlja isključivo na temelju interne strukture tekstualnog elementa.

Jedinice, desetice, brojevi od jedanaest do devetnaest, stotice, nazivi mjeseca u godini te nazivi posebnih simbola i znamenaka zapisani u odgovarajuće liste. Funkcije koriste te liste za gradnju odgovarajućeg niza riječi. Za jednoznamenkaste brojeve dovoljno je upariti naziv iz liste i znamenku. Složeni brojevi nastaju povezivanjem jednočlanih dijelova, primjerice, stotica, desetica i jedinica. Pri gradnji složenog broja nastoji se postići pravilni oblik riječi unutar broja pa će tako biti *dvije tisuće*, ali *pet tisuća*. Posebno su razvijene funkcije za pojedine članove složenog broja koje koriste pripadajuće liste definirane na početku modula i pozive ostalih funkcija. Ulančanim je pozivima na elegantan način riješen problem gradnje složenih brojeva. Pri trenutnom stupnju razvoja, ostvarena je normalizacija brojeva manjih od milijun, ali korištenjem razrađenog koncepta lako je proširiti podsustav tako da radi za po volji velike brojeve.

Očekivani oblici oznaka vremena su *hh:mm*, *hh:mm.ss* i *hh:mm:ss*. Za pretvorbu vremenskog zapisa u riječi korištene su funkcije za dvoznamenkaste brojeve i postupak za dodavanje vremenskih jedinica u pravilnom obliku. Kako su brojevi rodom prilagođeni vremenskim jedinicama, jedan od argumenata funkcijama za gradnju brojeva naznačuje koji oblik broja trebaju vratiti.

Nekoliko se primjera pretvorbe nalazi u nastavku:

02:02 → dva sata i dvije minute

13:30 → trinaest sati trideset minuta

01:32 → jedan sat trideset dvije minute

00:07 → ponoć i sedam minuta

23:05:06 → dvadest tri sata pet minuta šest sekundi

08:51.11 → osam sati pedeset jedna minuta jedanaest sekundi

Uz brojke često idu simboli poput slovnih oznaka novčane jedinice ili postotka. Tako primjerice, oznaka američkog dolara prethodi znamenkama zapisan broj, dok znak eura dolazi nakon posljednje

znamenke. Za prepoznavanje takvih slučajeva korišteni su regularni izrazi, a pretvorba je ponovno obavljena funkcijama za gradnju složenih brojeva i dodavanjem odgovarajuće riječi. Primjeri:

3% → tri posto

$235 → dvjesto trideset pet dolara

$21071 → dvadeset jedna tisuća sedamdeset jedan dolar

1€ → jedan euro

55¢ → pedeset pet centi

4.4.2. Miješani znakovni niz

Miješani znakovni niz može sadržavati slova, znamenke i ostale simbole, a ne pripada niti jednom od prethodno definiranih oblika. Uvođenjem miješanog znakovnog niza osigurana je pretvorba bilo kojeg teksta u riječi što je vrlo važno za interpretaciju tekstualnog izlaza osobnog računala. Na očekivan način normalizirat će se adrese webskih sjedišta, putanja do datoteka, nazivi datoteka, ali i bilo koja durga kombinacija znakova. Slovni podnizi pretvaraju se u cjelovite riječi, simobli se supstituiraju odgovarajućim nazivima, a znamenke se normaliziraju posebno. Primjeri prikazuju korak normalizacije:

www.x123.hr → www točka x jedan dva tri točka hr

adresa@pošta.com → adresa et pošta točka com

/root/home/datoteka_0 → kosa crta root kosa crta home kosa crta datoteka donja crta nula

Izrazima s desne strane izgovor pridružuje jezični modul.

Uočen je posebni oblik miješanog niza koji se prema pravili hrvatskog jezika ne smije izgovarati na ovaj način. To su polusloženice, primjerice: *spomen-ploča*, *čovjek-riba*. Među riječima je spojnica i ona se ne smije supstituirati odgovarajućim nazivom nego jednostavno ukloniti, a riječi odvojiti. Naime, riječi pisane spojnicom imaju dva naglaska.

4.5. Leksikon

Hrvatski jezik sadrži sistematičan odnos ortografskog zapisa i izgovora riječi. Zbog toga je moguće napisati skup pravila za tvorbu izgovora. Osim što je to u ovom slučaju jednostavnije od pripreme

strukturirane fonetske transkripcije s oznakom naglasaka za desetke tisuća riječi, korištenje pravila općenitiji je i istraživački zanimljiviji pristup. Dobro napisan skup pravila može u gotovo svim slučajevima osigurati ispravnu pretvorbu slova u foneme. Zato nije potrebno imati mnogo riječi u leksikonu, već samo pokriti iznimke, strane riječi i kratice.

S druge strane, za određivanje naglašenog sloga u hrvatskom jeziku ne postoji takav skup pravila koji će uvijek dati jednoznačni i točan odgovor. Nemoguće je napisati algoritam za nadopunjavanje izgovora ispravnom oznakom naglaska, iako se uz osobit trud učestalost pogrešnog naglašavanja može minimizirati. Korištenje velikog leksikona, koji sadrži cjeloviti izgovor uključujući naglaske, bolje je rješenje ovog problema.

Razmatranjem oba aspekta, kao najbolje rješenje čini se ugradnja velikog leksikona koji bi osigurao ispravne izgovore za što više riječi, ali i potpunog skupa pravila za riječi koje se ne nalaze u leksikonu. Unatoč tome, odlučeno je u ovom trenutku ipak ne ugrađivati mnogo zapisa u leksikon. Položaj sloga u riječi često nije ključan za općenitu razumljivost sustava, a primjenom određenih pravila, broj riječi s pogrešnim naglaskom može se umanjiti. Dodavanje novih zapisa u leksikon može se obaviti u bilo kojem trenutku kasnije razvoja.

Leksikon za hrvatski jezik izrađen je ručno, a sadrži sljedeće kategorije riječi:

1. Slova abecede; koriste se kada je tekstualni element akronim ili miješani znakovni niz s jednoslovnim podnizom. Primjeri zapisa:

```
(lex.add.entry '("p" nn (((p e) 0))))
(lex.add.entry '("r" nn (((e r) 0))))
(lex.add.entry '("s" nn (((e s) 0))))
(lex.add.entry '("š" nn (((e sx) 0))))
```

2. Jednoslovne oznake poput dolara, postotka, plusa itd. U slučaju da zbog neodgovarajućeg konteksta takve oznake nisu raspisane tijekom normalizacije teksta, one moraju biti adekvatno zamijenjene na ovom mjestu jer bi se pojavila pogreška prilikom provođenja pravila izgovora. Primjeri zapisa:

```
(lex.add.entry '("$" n (((d o) 0) ((l a r) 0))))
(lex.add.entry '("%" n (((p o) 1) ((s t o) 0))))
(lex.add.entry '("@" n (((e t) 0))))
```

3. Kratice koje nisu raspisane u riječi u postupku normalizacije teksta. Primjeri zapisa:

```
(lex.add.entry '("tzv" n (((t a ) 0) ((k o) 1) ((z v a) 0) ((n i) 0) )))
(lex.add.entry '("prof" n (((p r o) 1) ((f e) 0) ((s o r) 0))))
```

4. Iznimke za koje hrvatska pravila izgovora ne vrijede. To su strane riječi, međunarodni pojmovi, osobna imena, nazivi gradova, organizacija i slično. Primjeri zapisa:

```
(lex.add.entry '("mail" n (((m e j l) 1))))
(lex.add.entry '("skateboard" n (((s k e j t) 1) ((b o r d) 0))))
(lex.add.entry '("jack" n (((dx e k) 0))))
(lex.add.entry '("bruxelles" n (((b r i) 1) ((s e l) 0))))
(lex.add.entry '("deutsch" n (((d o j cx) 0)) ))
```

Za sve ostale riječi primjenjuju se standardna pravila za tvorbu izgovora.

4.6. Pravila za tvorbu izgovora

Postupak fonemske transkripcije za hrvatski jezik razmjerno je jednostavan. Svaki grafem ima svoj fonemski par. Ipak, u određenim sljedovima dolazi do promjena fonema, odnosno njihovog ujednačavanja koje olakšava rad govornih organa. Preslikavanja, kao i sve promjene, trebaju biti pokrivene skupom pravila za tvorbu izgovora.

Prije prelaska s grafema na foneme, ulaz je potrebno normalizirati što podrazumijeva pretvorbu svih slova u mala i preslikavanje digrafa u jedinstvene elemente. Normalizacijski je podskup pravila jednostavan jer pretvorbe ne ovise o kontekstu pojedinog grafema. Primjeri:

```
( [ ć ] = ć )     ( [ d ž ] = dž )
( [ Đ ] = đ )     ( [ D ] = d )
```

Iz normaliziranog ulaza sada se može obaviti preslikavanje grafema u foneme. Oznake fonema moraju odgovarati onima iz definicije skupa fonema. Primjeri:

```
( [ ć ] = cy )    ( [ d ž ] = dzx )
( [ đ ] = dy )    ( [ d ] = d )
```

Među tim pravilima su i ona za preslikavanje stranih grafema u hrvatske foneme poput:

```
( [ x ] = k s )   ( [ w ] = v )
( [ y ] = i )     ( [ q ] = k u )
```

Ostala pravila izgovora odnose se na stapanje fonema, jednačenje po zvučnosti, uklanjanje određenih suglasnika iz grupe glasova i slično. Ona se primjenjuju s obzirom na kontekst pojedinog fonema pa je zato prikladno definirati skupove koji će se koristiti za opis konteksta. Tako se u svrhu realizacije preostalih pravila definiraju sljedeći skupovi fonema:

- samoglasnici (A): *a e i o u*

- suglasnici (SG): *b c č ć d d dž đ f g h k j l lj m n nj p r s š t v z ž*
- zapornici (OC): *b p t d k g*
- zvučni suglasnici (ZV): *b d g dž đ z ž*
- bezvučni glasovi (BV): *p t k c č ć f s š h*
- suglasnici bez *v*, *j* i *r* (IZ): *b c č ć d d dž đ f g h k l lj m n nj p s š t z ž*

U tablici 1 nalaze se sistematizirana pravila fonemske transkirpcije za hrvatski jezik [12]. Situacije, u kojima se pravila primjenjuju, označene su plusom, a iznimke minusom. Katkad se primjenom jednog pravila fonemski niz promijeni tako da se na njemu mora provesti neko drugo pravilo. Na taj se način pravila ulančavaju, a redoslijed kojim se primjenjuju na pojedini primjer iz tablice, stoji u posljednjem stupcu. Pritom brojka označava redni broj sljedećeg ili prethodnog pravila u nizu, a zvjezdica se odnosi na trenutno pravilo.

Tablica 1: *Pravila izgovora za hrvatski jezik*

Broj	**Pravilo**	**Korištenje**		**Primjer**	
1	Uklanjanje suglasnika *t* i *d* iz grupa *st*, *zd* i *žd* ako nakon grupe slijedi suglasnik	+	posuđenice	*rostfraj* → /rosfraj̯/	
		+	imenice ženskog roda izvedene iz imenica muškog roda koje završavaju na *ist*	*rostfraj* → /feminiskińa/	
		+	na mjestima spoja u složenicama	*postdiplomski* → /posdiplomksi/	*2
		-	kada grupu slijedi suglasnik *v*, *j* ili *r*	*bratstvo* → /bratstvo/	*25
		-	grupa se nalazi u prvom slogu	*istkati* → /istkati/	
2	Zamjena niza suglasnika zvučnim, odnosno bezvučnim parovima ovisno o zvučnosti posljednjeg suglasnika	+	niz od dva suglasnika	*postdiplomski* → /pozdiplomksi/	1*
		+	niz od tri ili više suglasnika	*predstava* → /pretstava/	*5
		+	dva susjedna suglasnika su zvučno-bezvučni par	*subpolaran* → /suppolaran/	*4
		-	suglasnik *v* posljednji je u nizu	*bratstvo* → /bratstvo/	1*5
3	Dodavanje fonema /j̯/ između dva samoglasnika od kojih je barem jedan *i* ili *e*	+		*mie* → /mij̯e/	*5
4	Stapanje uzastopnih fonema u jedan fonem	+		*subpolaran* → /supolaran/	2*

5	Kominiranje dvaju ili više grafema u jedan fonem u određenim situacijama	+	*ije* → /ie/ (dvoglas)	*mie* → /mie/	3*
		+	*tc* → /c/	*bitci* → /bici/	
		+	*ts* → /c/	*bratstvo* → /bractvo/	12*
				predstava → /prectava/	2*
		+	*tč* → /č/	*mlatče* → /mlače/	
		+	*tć* → /ć/	*odćarlijati* → /oćarlijati/	2*
		+	*tš* → /č/	*predškolski* → /prečkolski/	2*
		+	*dz* → /c/		2*
		+	*ddž* → /ʒ/	*sladoleddžija* → /sladoleʒija/	
		+	*dđ* → /ʒ́/	*poddđakon* → /poʒ́akon/	
		+	*sš* → /š/	*uzšetati* → /ušetati/	2*
		+	*zž* → /ž/	*razžvakati* → /ražvakati/	
		+	*sč* → /šč/	*rasčlaniti* → /raščlaniti/	
		+	*sć* → /šć/		
		+	*zdž* → /žʒ/		
		+	*zđ* → /žʒ́/	*razđakoniti* → /ražʒakoniti/	
		+	*np* → /mp/	*jedanput* → /jedamput/	
		+	*nb* → /mp/	*stanben* → /stanben/	
		+	*nm* → /m/		
		+	*cč* → /č/		
		+	*cć* → /ć/		
6	Pretvorba suglasnika *r* u slogotvorno /r̩/	+	suglasnik *r* je na početku riječi, a slijedi ga suglasnik	*rzati* → /r̩zati/	
		+	suglasnik *r* je na kraju riječi, a prethodi ga suglasnik	*žanr* → /žanr̩/	
		+	suglasnik *r* je između dva suglasnika	*prst* → /pr̩st/	

Sva su pravila, osim prvog, u potpunosti ugrađena u podsustav za tvorbu izgovora.

Fonemski niz potrebno je podijeliti u slogove. U tu svrhu je napisan algoritam koji detektira samoglasnike i slogotvorni suglasnik /r̩/ kako bi na temelju toga odredio granice slogova. Jednosložne i dvosložne riječi dobivaju naglasak na prvom slogu, dok za višesložne riječi problem određivanja mjesta naglaska nije riješen. Zasad se naglasak uvijek postavlja na drugi slog. Nekoliko najčešće korištenih višesložnih riječi, kod kojih je naglašavanje drugog sloga nepravilno, ručno je upisano u leksikon.

5. Mjerenje kvalitete sintetiziranog govora

Kvaliteta sustava za umjetnu tvorbu govora odnosi se na razumljivost i prirodnost sintetiziranog govora. Pritom se ne smije isključiti prikladnost u konkretnoj primjeni. Tako, primjerice, čitači zaslona za slijepe osobe moraju posjedovati visoku razumljivost na većim brzinama čitanja, dok je prirodnost izgovora manje bitna. S druge strane, u multimedijskim primjenama izražajnost i prirodnost imaju presudnu ulogu.

Razumljivost i prirodnost sintetiziranog govora može se odrediti iz funkcionalnih i prosudbenih testova izlaznog govora na akustičkoj razini. Za takve su testove potrebni ispitanici koji izražavaju subjektivno mišljenje o onome što su čuli – oni bilježe percipirani sadržaj reproduciranog govora ili ispunjavaju različite ankete s pitanjima o razumljivosti, ugodnosti i prirodnosti govora te mogućoj uporabi sustava. Sistematično provođenje mjerenja kvalitete način je dobivanja odgovora mogu li osobe nenaviknute na sintetizirani govor dovoljan udio sadržaja pravilno razumjeti i smatraju li da je prirodnost izgovora zadovoljavajuća u određenom kontekstu. Na temelju rezultata, sustav se može dalje usavršavati i nadograđivati.

Kada se mjerenja za različite sustave provode istom metodologijom u istim ili sličnim uvjetima, ona mogu poslužiti za usporedbu sintetizatora govora. Mogućnost usporedbe bitna je pogotovo prilikom istraživanja i razvoja novih pristupa i metoda.

Neriješena pitanja kod akustičkih testova obuhvaćaju izbor ispitanika, sadržaj testova i procedure testiranja. Vrlo je teško ili nemoguće ustvrditi koja metodologija ispitivanja kvalitete daje točne rezultate.

Kod sustava za pretvorbu teksta u govor, osim akustičkih karakteristika, važni su jezični aspekti. Tako se za različite ulazne tekstovne elemente provjerava rad podsustava za normalizaciju teksta i fonemsku transkripciju. Često se za ispitivanja točnosti ortografsko-fonemske pretvorbe koriste članci iz novina i časopisa, nazivi gradova i zemalja te osobna imena.

U sklopu ovoga rada provedeno je akustičko ispitivanje sustava za sintezu govora. Specifični je cilj mjerenja dobiti ocjenu kvalitete izrađenog sustava za hrvatski jezik. Kako su se koristile standardne metode prihvaćene od strane drugih istraživača, rezultati se mogu usporediti sa sličnim sustavima. Također, mjerenjem se može odrediti uspješnost sinteze za pojedine glasove i identificirati problematične hrvatske foneme.

5.1. Akustičko ispitivanje

Prilikom akustičkih ispitivanja, ispitanici slušaju dijelove sintetiziranog govora i odgovaraju na pitanja. Često trebaju napisati koje su foneme, riječi ili rečenice čuli. Ispitni materijali uglavnom su usmjereni na suglasnike jer samoglasnici nisu problematični za sintezu.

Ponavljanje postupka ispitivanja na istoj grupi ispitanika može rezultirati poboljšanjem rezultata jer se osobe privikavaju na sintetizirani govor i lakše ga razumiju nakon svake iteracije testova. S druge strane, tijekom vremena opada koncentracija ispitanika, osobito kod segmentnih ispitivanja. To je važno imati na umu kod izbora ispitanika i planiranja postupka ispitivanja.

Ispitivanje se može provesti na različitim razinama ovisno o tome koju informaciju želimo njime dobiti [13]. Tako se, pirmjerice, segmentna ispitivanja provode na kratkim dijelova govora kako bi se ustanovila njihova razumljivost, dok se za ispitivanja na razini rečenica koriste stvarne riječi poredane u rečenice kako bi se ocijenila pravilnost shvaćanja sintetiziranog govora.

5.1.1. Segmentna ispitivanja

Kada su akustička ispitivanja usmjerena na pojedine isječke govora, radi se o segmentnim ispitivanjima. Općenito vrijedi da je uz dobro raspoznavanje segmenata lako raspoznati riječ, koja se od njih sastoji, neovisno o intonaciji i trajanjima segmenata [14].

Ovaj je pristup za mjerenje kvalitete govora zanimljiv istraživačima jer je:

- dobra segmentna kvaliteta preduvjet dobre ukupne kvalitete
- definiran kriterij procjene kvalitete – raspoznavanje fonema
- postupak inherentno precizan – fonem je ili ispravno ili pogrešno prepoznat
- često moguće izraditi testove tako da postupak testiranja bude jednostavan i ispitanike ne treba dodatno pripremati za ispitivanje
- često dovoljan relativno mali broj ispitanika

Postoji više različitih metoda segmentnih ispitivanja. U ovom je istraživanju provedeno standardno segmentno ispitivanje razvijeno u sklopu projekta *ESPRIT* grupe *Speech Assessment Methods.* Postupak je relativno kratkotrajan i daje precizne informacije o načinu na koji ispitanici percipiraju suglasnike u sintetiziranom govoru. Za gradnju testova koristi se strukture CV, VC i VCV pri čemu

C označava suglasnik, a V samoglasnik. Strukture su odabrane na način da svaka sadrži po jedan suglasnik i da pokriju slučajeve različitih pozicija suglasnika. Samoglasici se ne ispituju, već samo služe za pružanje različitog konteksta suglasnicima. Tipično se koriste samoglasnici /a/, /i/ i /u/. Primjeri stavaka za testiranje su: *pa*, *ap*, *apa*, *ki*, *ik*, *iki*. Pitanja su otvorenog tipa pa ispitanici nakon svakog preslušanog izgovora biraju jedan iz skupa svih suglasnika. Razmak između dvije reprodukcije uvijek je četiri sekunde. Jedan test sastoji se od svih suglasnika, na svim pozicijama i u sva tri samoglasnička konteksta, što je za hrvatski jezik ukupno 225 stavaka. Stavke su poredane slučajnim redoslijedom.

5.1.2. Ispitivanja na razini rečenica

Za ispitivanje na razini rečenica koriste se skupovi rečenica koje se često slažu tako da pojava riječi u njima modelira učestalost riječi u tom jeziku. Svrha ovog pristupa je izmjeriti koliko točno slušatelj shvaća sintetizirani govor. Naime, iako se kratki dijelovi govora mogu propustiti ili krivo razumjeti, konačan odgovor ispitanika i dalje može biti ispravan.

Postoji više metoda ispitivanja na razini rečenica. Za ovo je istraživanje odabrana metoda semantički nepredvidljivih rečenica. Pri generiranju ispitnih rečenica koristi se pet uobičajenih sintaktičkih struktura. Riječi se uzimaju slučajnim izabirom pa rezultantna rečenica nije smislena kao cjelina. Osobine metode semantički nepredvidljivih rečenica:

- lako rukovođenje ispitivanjem – nije potrebna priprema ispitanika
- odgovori se mogu automatski vrednovati
- usporedivost za različite jezike
- odgovori su otvorenog tipa

5.2. Ispitanici i metode

Na ispitivanju je sudjelovalo šestero izvornih govornika hrvatskog jezika. Nitko od ispitanika nije imao problema sa sluhom, nije bio posebno pripreman za testiranje niti je ranije dulje vrijeme koristio sustave za umjetnu tvorbu govora.

Prije akustičkog ispitivanja, zbog podešavanja opreme za reprodukciju zvuka i privikavanja na sintetizirani govor, ispitanici su poslušali nekoliko sintetiziranih rečenica. Nakon toga su obavljena

dva standardna segementna testa i jedan test semantički nepredvidljivih rečenica.

Ispitivanje je obavljeno u laboratorijskom prostoru, svaki ispitanik koristio je po jedno osobno računalo i slušalice. Testove su rješavali istovremeno, a odgovore bilježili u posebno razvijeno korisničko sučelje izrađeno u *Matlabu* samo za ovo ispitivanje. Ukupno trajanje, koje uključuje objašnjavanje procedure, podešavanje opreme, uvodno slušanje u svrhu privikavanja, rješavanja testova i kratkih pauza među njima, bilo je oko sat vremena.

5.2.1. Standardni segmentni test

Kako je razmak između dvije stavke standardnog segmentnog testa četiri sekunde, ukupno trajanje jednog testa je oko 15 minuta. Tijekom ispitivanja provedena su dva testa s različitim redoslijedom stavki.

U hrvatskom jeziku svaki fonem ima svoj grafemski par pa je bilo jednostavno odabrati način bilježenja odgovora. Ispitanici su jednostavno trebali upisati suglasnik koji su čuli u sintetiziranom segmentu govora. Za provođenje standardnih segmentnih testova pripremljena je *Matlab* skripta koja reproducira zvučni zapis sa segmentima sintetiziranog govora i dopušta unos podataka s tipkovnice. Ispitanici u svaki redak upisuju po jedan suglasnik. Skripta ima mogućnost odmah obraditi rezultate i za svaki suglasnik izračunati postotak točnih odgovora, ali može i pohraniti sve upisane odgovore.

5.2.2. Semantički nepredvidljive rečenice

Za ispitivanje na razini rečenica koristilo se po pet rečenica za svaku od pet rečeničnih struktura. Pauza između dvije rečenice iznosilo je 15 sekundi kako bi ispitanici stigli zabilježiti odgovor. Ukupno je trajanje testa bilo oko osam minuta.

U ovom su testu korištene su sljedeće rečenice:

Skupina 1

Kotač raste na gradskoj osmici.
Vulkan klima u premalom šeširu.
Ruka odlazi na veliku stranu.
Škola spava sa svojom kosom.
Zumbul trubi u osami.

Skupina 2

Rozi prijatelj jede loptu.
Grmolika podmornica ljulja konzervu.
Prugasti glumac množi biljke.
Žustra pegla prodaje inje.
Crveni sir bježi od luka.

Skupina 3

Dolazi klobučar ili kiklop.
Bojim se satova i satira.
Kombiniraj trčanje i testiranje.
Volim turbine i krpelje.
Pomiče snijeg ili pjenu.

Skupina 4

Kada morž piše crni zid?
Zašto glina umanjuje adresu?
Koje stablo truje tuniku?
Gdje filozof kažnjava cipelu?
Zašto nevidljivost želi prijatelja?

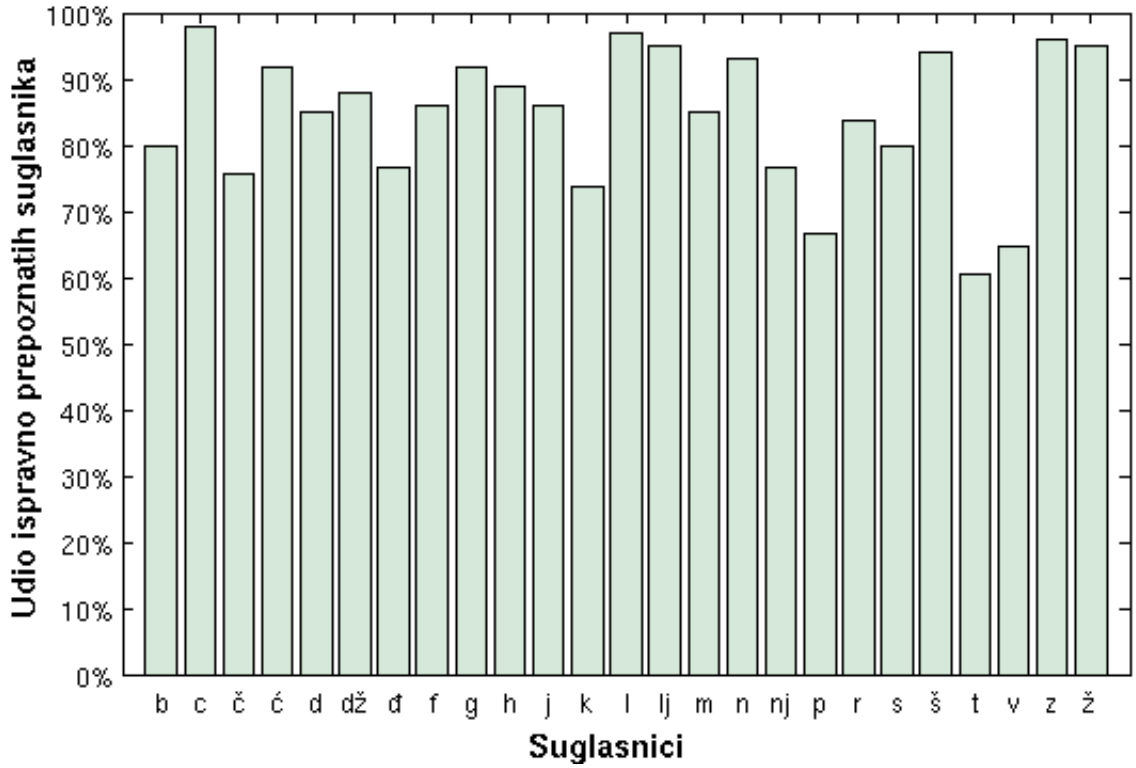
Skupina 5

Račun je pomogao rastućem brodu.
Mramor je nahranio oronulog pijetla.
Mrkva je prestigla zlovoljnog brata.
Tratinčica je nadglasala sedam pelikana.
Sloboda je prerasla zelenu tugu.

5.3. Rezultati

Pri ocjeni segmentalnih testova za svaki je suglasnik određen udio ispravno prepoznatih glasova u ukupnom broju pojavljivanja u testu. Postotci su izračunati za svaki test i za svakog ispitanika posebno, a zatim usrednjeni u konačne rezultate. Za obradu rezultata napisana je *Matlab* skripta, a svi su odgovori i međurezultati pohranjeni te se mogu koristit za složenije obradbe.

Ukupni udio točno prepoznatih glasova za sve testove i sve ispitanike iznosi 84.3%. Rezultati za pojedine suglasnike prikazane su slikom 3.



Slika 3: *Rezultati standardnog segmentnog ispitivanja.*

Rezultati ispitivanja na razini rečenica govore koliki je udio ispravno prepoznatih riječi u cijelom ispitnom materijalu. Obradbom odgovara svih sudionika ispitivanja, dobiveno je 96.1% točno prepoznatih riječi.

Tablicom 2 prikazani su rezultati oba ispitivanja po ispitanicima.

Tablica 2: *Rezultati po ispitanicima*

Ispitanik	**Rezultati standardnog segmentnog ispitivanja**	**Rezultati ispitivanja semantički nepredvidljivim rečenicama**
1	87.7%	96.4%
2	84.7%	94.6%
3	82.4%	95.5%
4	83.1%	94.6%
5	81.7%	98.2%
6	87.0%	97.3%

Standardna devijacija segmentnog ispitivanja je 2.5%, a ispitivanja na razini rečenica 1.5%.

5.4. Rasprava

Ispitivanje je pokazalo da je postignuta zadovoljavajuću razumljivost sintetiziranog govora. Iako su rezultati ohrabrujući, oni također upućuju na prostor za moguća poboljšanja. Popravljanjem granica fonema i ugradnjom kvalitetnijeg modela trajanja, razumljivost se može značajno popraviti. Koristeći ove rezultate moguće je efikasno isplanirati budući razvoj.

Segmentnim testom pokazano da je foneme /t/, /v/, /p/ i /k/ najteže pravilno prepoznati. Učestalosti pravilnog prepoznavanja za te foneme redom su 60.6%, 64.6%, 66.7% i 73.7% što je znatno niže od prosjeka za sve suglasnike. Najviše se pogrešaka odnosi na zamjenu glasa njihovim zvučnim, odnosno bezvučnim parnjakom. Gotovo svi pogrešni odgovori za glas /t/ glasili su /d/.

Isti su se fonemi pokazali problematičnim i pri ispitivanju na razini rečenica. Tako je najčešća pogreška bila zamjena riječi *piše* riječju *diše*. Iz rezultata se ipak može zaključiti da loša učestalost prepoznavanja kod samo nekoliko fonema ne može značajnije ugroziti ukupnu razinu shvaćanja sintetiziranog govora. Zanimljivo je da je ispitanik s najlošijim rezultatima segmentnog testa imao najveći broj točno prepoznatih riječi u ispitivanju na razini rečenica.

Tablicom 3 prikazani su rezultati sličnih ispitivanja razumljivosti za druge suvremene difonske sintetizatore. Ispitivanje prepoznavanjem izoliranih riječi dovoljno je slično metodi semantički nepredvidljivih rečenica da rezultati mogu poslužiti za usporedbu.

Tablica 3: *Rezultati ispitivanja kvalitete nekih sintetizatora govora*

Jezik	**Značajke sustava za sintezu**	**Opis ispitivanja**	**Rezultati**
singaleški	Festival, difonska sinteza, 2007. [15]	prepoznavanje izoliranih riječi	72%
hrvatski	difonska sinteza, 2008. [2]	semantički predvidljive rečenice (preuzete iz vremenske prognoze)	76%
arapski	Festival, difonska sinteza, 2005. [16]	prepoznavanje izoliranih riječi	85%
švedski	difonska sinteza, 1989.	segmentni test, VCV strukture	87%
poljski	Festival, difonska sinteza, 1998. [3]	standardni segmentni test, CV, VC i VCV strukture te suglasničke grupe	96%
		semantički nepredvidljive rečenice	99%

Standardno segmentno ispitivanje ne uključuje samoglasnike. Također nisu pokriveni slučajevi suglasnika kada se pojavljuju u zvučnim, odnosno bezvučnim grupama što bi za difonsku sintezu moglo biti od velike važnosti. Promatrano iz aspekta baze segmenata, ovim testovima nisu obuhvaćeni razredi difona CC niti VV pa se za buduća istraživanja predlaže dodavanje tih struktura.

Dvjema metodama, provedenim u sklopu ovog istražvinja, pokazano je da, iako obje metode mjere razumljivost sintetiziranog govora, one se zapravo odnose na različite aspekte razumljivosti i polučuju različitim rezultatima. Zbog toga, prilikom mjerenja kvalitete treba kombinirati veći broj različitih metoda. Na taj se način može dobiti cjelovitija slika kvalitete sustava za sintezu govora.

6. Zaključak

Iako hrvatski jezik sadrži razmjerno mali skup fonema i ima jasna pravila ortografsko-fonemske pretvorbe, ipak postoje specifičnosti koji otežavaju izradu sustava za pretvorbu teksta u govor. Normalizacija teksta iznimno je složena jer raspisivanje brojeva i kratica podrazumijeva određivanje roda, broja i padeža riječi. Kod tvorbe izgovora, problem je detekcija naglasnog sloga koja se algoritamski teško može riješiti.

Izazovi također proizlaze iz činjenice da je ovo prvi pokušaj izrade cjelovitog sustava za pretvorbu teksta u govor. Čak i problemi, koji za hrvatski jezik nisu ništa složeniji u usporedbi s drugim jezicima, zahtijevaju izvorna rješenja u različitim stupnjevima razvoja. To nije samo slučaj s tekstualnim i jezičnim modulima, nego i s izborom skupa fonema, generiranjem liste difona, pripremom difonske baze, snimanjem i označavanjem sadržaja.

Ispitivanje kvalitete konzistentan je način dobivanja ocjene koja govori o razumljivosti sintetiziranog govora. Osim toga, rezultati ispitivanja važni su za planiranje daljnjeg razvoja, usporedbe tehnologija, mogućnosti primjene sustava, ali i procjenu relevantnosti samih metoda ispitivanja.

Vrijednost ovog sustava prije svega leži u funkcionalnom sustavu za pretvorbu teksta u govor koji se može koristiti u različitim primjenama. Kvaliteta je sintetiziranog govora zadovoljavajuća što je pokazano akustičkim ispitivanjima. Važnost ovog sustava također leži i u značajnom olakšanju razvoja novih hrvatskih sintetizatora i ugradnje novih govorničkih glasova. Automatsko označavanje fonema može se provesti bez preslikavanja između različitih jezika što će postupak učiniti uspješnijim.

Segmentno ispitivanje i ispitivanje semantički nepredvidljivim rečenicama metode su koje nikad ranije nisu korištene za ispitivanje kvalitete hrvatskih sintetizatora govora. Rezultati ovog istraživanja moći će poslužiti kao usporedba pri budućim ispitivanjima.

Nastavak razvoja sustava pretvorbe teksta u govor obuhvaća razvoj intonacijskih modela i modela trajanja, usavršavanje normalizacijskog modula i popravljanja granica fonema.

Zahvale

Zahvaljujem mentoru prof. dr. sc. Davoru Petrinoviću na brojnim korisnim savjetima, uloženom trudu i konstantnoj podršci.

Hvala svim sudionicima akustičkog ispitivanja na izdvojenom vremenu.

Popis literature

[1] A. Black, P. A. Taylor, “Festival speech synthesis system: system documentation (1.1.1)”, Human Communication Research Centre Technical Report HCRC/TR-83, 1997.

[2] M. Pobar, S. Martinčić-Ipšić, I. Ipšić, “Text-to-speech synthesis: a prototype system for croatian language”, *Engineering Review*, Vol.28 No.2, 2008., pp. 31-44

[3] D. Oliver, “Polish text-to-speech synthesis”, M.Sc. thesis, University of Edinburgh, 1998.

[4] J. Gros, N. Pavešić, F. Mihelič, “Text-to-speech synthesis: a complete system for the Slovenian language”, *Journal of Computing and Information Technology* – CIT 5, 1997., pp. 11-19

[5] K. Yoon, “Building a prosodically sensitive diphone database for a korean text-to-speech synthesis system”, Ohio State University, 2005.

[6] M. Wolters, “A diphone-based text-to-speech system for Scottish Gaelic”, University of Bonn, 1997.

[7] B. Möbius, R. Sproat, J. P. H. van Santen, J. P. Olive , "The Bell Labs German text-to-speech system: An overview", Proceedings of the European Conference on Speech Communication and Technology vol. 5: 2443-2446., 1997.

[8] A. Black, K. Lenzo, “Multilingual text-to-speech synthesis”, ICASSP, Montreal, Canada, 2004.

[9] A. Black, K. Lenzo, “Building synthetic voices”, Festvox system documentation, 2007.

[10] E. Barić et al., “Hrvatska gramatika”, Školska knjiga, 1997.

[11] A. Black, K. Lenzo, “Diphone collection and synthesis”, ICSLP2000, Beijing, China, 2000.

[12] B. Dropuljić, D. Petrinović, "Development of acoustic model for Croatian language using HTK", Automatika : časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije (0005-1144) 51 , 2010., pp. 79-88

[13] Sammy Lememtty, "Review of speech synthesis technology", Master's thesis, Helsinky University of Technology, 1999.

[14] C. Benoit, M. Grice, and V. Hazan, “The SUS test: a method for the assessment of text-to-speech synthe- sis intelligibility using semantically unpredictable sentences”, *Speech Communication*, vol. 18, 1996., pp. 381– 392

[15] R. Weerasinghe, A. Wasala, V. Welgama, K. Gamage, "Festival-si: a Sinhala text-to-speech system", University of Colombo School of Computing, 2008.

[16] M. Assaf, "A prototype of an Arabic diphone speech synthesizer in Festival", Master's thesis, Uppsala University, 2005.

Sažetak

Naslov: *Specifičnosti sustava za pretvorbu teksta u govor za hrvatski jezik*

Autor: Gordan Kreković

Ovaj rad opisuje prvi pokušaj izrade cjelovitog sustava za pretvrobu teksta u govor za hrvatski jezik. Sustav se temelji na difonskoj sintezi i razvijen je pomoću okruženja *Festival*. U svrhu prikupljanja ciljnih difona, generirano je 885 logatoma i snimljeni su njihovi prirodni izgovori. Granice fonema postavljene su postupcima automatskog označavanja, a zatim ručno korigirane. Razvijeni su i opisani moduli za normalizaciju teksta i jezičnu obradbu. Istaknute su specifičnosti hrvatskog jezika i izvorna rješenja primijenjena u našem sustavu. Provedeno je mjerenje kvalitete standardnim segmentnim ispitivanjem i semantički nepredvidljivim rečenicama. Rezultati su pokazali visoku razumljivost sintetiziranog govora – za segmente, prosječna učestalost pravilno prepoznatih suglasnika bila je 84.3%, a udio točnih riječi prilikom ispitivanja na razini rečenica 96.1%.

Ključne riječi: *sinteza govora*, *izrada difonske baze, normalizacija teksta, ispitivanje kvalitete govora*

Summary

Title: *Text-to-speech synthesis for the Croatian language*

Author: Gordan Kreković

This paper describes the first attempt at building a complete diphone based text-to-speech system for the Croatian language. The system was developed using the *Festival* environment. To collect all the target diphones, we generated 885 logatoms and recorded their natural pronunciation. Recordings were labeled automatically and the phone boundary marks were later corrected manually. Modules for text normalization and linguistic processing were developed and discussed. Specifics of the Croatian languages and our original solutions were presented. We conducted a speech quality evaluation using Standard Segmental Test and Semantically Unpredictable Sentences. The evaluation yielded a score of 84.3% for the segmental test and encouraging 96.1% for the test at the sentence level.

Keywords: *speech synthesis, diphone database preparation, text normalization, speech quality evaluation*

Životopis

Student sam druge godine diplomskog studija (smjer Informacijska i komunikacijska tehnologija, profil Obradba informacija) na Fakultetu elektrotehnike i računarstva, Sveučilišta u Zagrebu.

Tijekom osnovne i srednje škole, sudjelovao sam na natjecanjima iz informatike. Najzapaženiji uspjesi su oni na međunarodnom timskom natjecanju ACSL održanom u Sjedinjenim Američkim Državama i Kanadi – osvojeno drugo, a dvije godine kasnije i prvo mjesto. Dobio sam izravan upis na Fakultet elektrotehnike i računarstva. Preddiplomski studij (modul Računalno inženjerstvo) završio sam 2008. s prosjekom 5.00 i pritom stekao akademski stupanj sveučilišnog prvostupnika inženjera računarstva.

Bio sam voditelj nekoliko studentskih projekata. Najznačajnije iskustvo svakako je bilo vođenje međunarodnog raspodijeljenog studentskog tima (pet članova iz četiriju zemalja) na projektu "NRTRDE Processing System" u sklopu kolegija održanog u suradnji sa Sveučilištem Mälardalen u Švedskoj.

U matematičkoj XV. gimanziji održao sam niz predavanja za pripremu mladih informatičara za algoritamska natjecanja.

Dva sam puta dobio nagradu Josip Lončar koju Fakultet elektrotehnike i računarstva dodjeljuje najuspješnijim studentima u generaciji. Također sam stipendist Grada Zagreba.