



Samodejno tvorjenje govora iz besedil

Jerneja Gros

Dr. Jerneja Gros je raziskovalna sodelavka na Fakulteti za elektrotehniko in računalništvo Univerze v Ljubljani. Ukvarja se z raziskovalnim delom na področju govornih tehnologij, s posebnim poudarkom na postopkih za samodejno tvorjenje govora iz besedil in raziskovanjem proizvodnje govornenega jezika.

Linguistica et philologica

Jerneja Gros

Samodejno tvorjenje govora iz besedil

Postopek za izdelavo sintetizatorja slovenskega govora

© 2000, Založba ZRC, ZRC SAZU, Ljubljana

Izdal Inštitut za slovenski jezik Frana Ramovša ZRC SAZU

Za izdajatelja Varja Cvetko Orešnik

Založil Založba ZRC, (ZRC SAZU)

Za založnika Oto Luthar

Urednik

založništva Vojislav Likar

Oblikovanje zbirke Milojka Žalik Huzjan

Prelom Brane Vidmar

Tisk Littera picta, Ljubljana

Knjiga je izšla s podporo Ministrstva za znanost in tehnologijo
R. Slovenije.

Digitalna verzija (pdf) je pod pogoji licence CC BY-NC-ND 4.0 prosto
dostopna: <https://doi.org/10.3986/9616358219>.

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

811.163.6'34:681.3

681.3:811.163.6'34

811.163.6'322.6

GROS, Jerneja

Samodejno tvorjenje govora iz besedil : postopek za izdelavo
sintetizatorja slovenskega govora / Jerneja Gros. - Ljubljana : Založba
ZRC, ZRC SAZU, 2000. - (Linguistica et philologica)

ISBN 961-6358-21-9

109486080

JERNEJA GROS

SAMODEJNO TVORJENJE GOVORA IZ BESEDIL

POSTOPEK ZA IZDELAVO SINTETIZATORJA SLOVENSKEGA GOVORA

Ljubljana 2000

Z A L  Ž B A
Z R C

PREDGOVOR

Monografija *Samodejno tvorjenje govora iz besedil* je predelana doktorska disertacija, ki je bila izvedena pod mentorstvom prof. dr. Nikole Pavešiča in somentorstvom prof. dr. h. c. dr. h. c. Ludvika Gyergyeka, na Fakulteti za elektrotehniko v Ljubljani. K predelavi disertacije v monografijo sta me spodbudila akad. prof. dr. Janez Orešnik s Filozofske fakultete v Ljubljani in prof. dr. Varja Cvetko Orešnik z Inštituta za slovenski jezik Frana Ramovša. Prof. dr. Varja Cvetko Orešnik je predlagala, da se besedilo izda v obliki monografije v sklopu zbirke *Linguistica et Philologica*. Prvotno besedilo doktorske disertacije sem predelala in dopolnila z nekaterimi novostmi. Strogo tehnične dele disertacije sem zamenjala s krajšimi opisi. Prav tako sem izpustila mnogo številčnih podatkov o vrednostih prozodičnih parametrov v slovenskem govoru. Zainteresirani bralec jih lahko poišče v doktorski disertaciji v poglavjih, na katera se sklicujem v monografiji.

Mentorju in somentorju se lepo zahvaljujem za umerjanje in vodenje pri mojem raziskovalnem delu. Najlepša zahvala gre tudi prof. dr. Francetu Miheliču, mojemu neuradnemu delovnemu mentorju, saj je neprestano bedel na mojim delom, ga usmerjal in komentiral. Še posebej se želim zahvaliti prof. dr. Tatjani Srebot Rejec. Njene razlage meritev prozodičnih parametrov slovenskega govora so me vodile pri modeliranju tonemskega naglasa in pri modeliranju trajanja glasov.

Zahvaljujem se tudi nadvse pridni in prizadevni skupini diplomantov – Ermini, Mariji, Alešu, Andreju, Boštjanu, Mateju, Milanu, Robertu, Tomažu, Urošu in Viktorju, ki so bili vedno pripravljene udejaniti nove zamisli in izboljšave na sintetizatorju govora. Brez njihove pomoči izvedba tako obsežnega dela ne bi bila možna.

Za pomoč pri delu se zahvaljujem tudi sodelavcem iz Laboratorija za umetno zaznavanje, predvsem mag. Simonu Dobrišku za pomoč pri označevanju govornih signalov in dr. Ivi Ipšiču za poskus samodejne segmentacije difonske baze. Najlepša hvala mag. Mariu Žgancu, ki je pomagal pri izdelavi modula za grafemsko-fonetično pretvorbo besedila in pri definiciji modela za nastavljanje trajanja glasov. Dr. Tomažu Erjavcu hvala za govorni korpus posnetkov radijskega napovedovalca, ki smo ga uporabili kot referenčni govor pri preizkusu sintetizatorja govora.

Jerneja Gros

Ljubljana, september 2000

1. UVOD

Človeški govor se uvršča med najstarejše in najbolj uporabljane načine sporazumevanja med ljudmi. Zato ne preseneča, da so ljudje pogosto poskušali graditi stroje, ki bi bili sposobni samodejno obdelovati govor, bodisi da bi ga prepoznali ali pa obratno, sintetizirali.

V štiridesetih letih tega stoletja je bil predstavljen prvi električni stroj, ki je samodejno pretvarjal besedila v govor (Dudley Voder 1939, Bell Labs) [Flanagan 72, str. 204–210]. Po začetnem zanosu so razvijalci kmalu spoznali omejitve, ki jih je postavljalo takratno stanje tehnike, in uvideli, da ne morejo umetno sintetizirati kakovostnega govora [Pierrel 87].

V zadnjem času se je s hitrim razvojem računalniške tehnologije ustvarila možnost za govorno sporazumevanje med ljudmi in stroji. Ti sistemi morajo biti sposobni tako prepoznavati in razumeti govorne informacije kakor tudi tvoriti oz. sintetizirati govor za posredovanje povratne informacije v okolje.

Sinteza govora je proces umetnega ustvarjanja govora, ki ga lahko uporabljamo za najrazličnejše namene [Dutoit 93a, str. 19–20].

Sistemi za pretvorbo besedila v govor omogočajo posredovanje informacij, kot so teletekst, elektronska pošta, faksi in druga besedila, brez potrebe po njihovi vidni predstavitvi, kar je še zlasti uporabno za slepe in slabovidne osebe. Informacije lahko tako posredujemo tudi tam, kjer to iz različnih razlogov do sedaj ni bilo možno, npr. v nevarnem delovnem okolju ali 24-urno zagotavljanje svežih informacij.

Razvoj kakovostnega sintetizatorja govora ni pogojen le s tehnologijo samega tvorjenja govornega signala, temveč tudi z osnovnimi raziskavami govora in jezika. Ko je bila dosežena dovolj velika kakovost izgovora posameznih govornih segmentov, se je pokazala potreba po upoštevanju nadaljnjih jezikoslovnih nivojev. Vključevanje novih znanj je potekalo postopno. Najprej so bila uporabljena znanja iz fonetike, nato iz skladnje in končno tudi delno iz semantike.

Samodejno tvorjenje govora iz besedil se uvršča v ožje raziskovalno področje sporazumevanja med človekom in strojem [Furui 89, str. 313], [Pavešić 92, str. 1]. Načrtovanje govornega vmesnika med človekom in strojem vključuje znanje iz različnih vej elektrotehnike, informatike in jezikoslovja, kot so [Rabiner 93] glasoslovje, oblikoslovje, psihoakustika, obdelava naravnega jezika, teorija informacij, teorija obdelave signalov in umetna inteligenca.

1.1 Psihološke in fiziološke osnove govora

Že bežni opis mehanizma govorjenega branja pri človeku nakazuje težavnost avtomatizacije tega postopka.

Človeški govor je zapleten psihofizičen proces, ki se razvija pod vplivom bioloških dejavnikov in socialnega okolja [Omerza 70, str. 6]. Govorec nam oblikovano stvarnost svoje zavesti in podzavesti ubeseduje, tj. daje ji izrazno glasovno ali pisno podobo [Toporišič 69, str. 595]. Oblikovanost zavesti pomeni enkratno duševno zajetje dela stvarnosti. Besedilo je eksistenčna oblika resnične ali umišljene stvarnosti, pretvorjene v besedo. Ima slušno in vidno podobo, pred t. i. izrekom pa nevrofiziološko [Toporišič 69, str. 616]. Najmanjše besedilo ima obliko enostavčne povedi, tj. besed, zbranih okoli osebne glagolske oblike.

Postopek ustvarjanja slušne podobe besedila sestavlja množica zaporednih procesov. Začetnemu stanju priprave koncepta sledijo besedni izbor, fonološko kodiranje, fonetično kodiranje in izgovor besedila [Levelt 96]. Proženje ustreznega koncepta, ki posebej namen, ki ga želimo izraziti, lahko predstavimo z računalniškim modelom, kjer se v semantični mreži konceptov aktivira želeni koncept. Sledi besedni izbor, kjer se izmed množice lem, ki semantično ustrezajo izbranemu konceptu, izbere primerna. Izbor leme sproži vzpostavitev povezave s fonološko kodo besede. Dostopni čas do fonološke kode je odvisen od pogostosti besede, ki jo lema ponazarja [Levelt 96]. Fonološko kodiranje zajema določanje segmentnih in metričnih lastnosti besede. Metrične strukture sosednjih besed se lahko združijo v večje celote, t. i. fonološke besede, ki so razgrajene po zlogih. Sledi fonetično kodiranje, kjer se za vsak fonološki zlog sproti preračuna izrazna oblika govoril. Tu se uporabi zlogovna zbirka, ki zajema izrazne oblike govoril za najbolj pogoste zloge. Takoj zatem se lahko začne izreka besedila.

Z nevrofiziološkega stališča poteka omenjeni postopek na naslednji način [Omerza 70, str. 8]. Slišana ali videna in spoznana beseda vzbudi v najrazličnejših predstavnih središčih v možganski skorji po asociacijskih zvezah predstavo ali misel, ki odloči reakcijo na to pobudo. Vsebina odgovora je skupek logičnega mišljenja. Z vprašanjem izzvani dražljaj se prenese iz možganske skorje v Brocovo središče, kjer so shranjeni motorični vtisi besed. Od tu potuje pobuda skozi motorično sfero in po motoričnem živcu do obkrajnih govornih organov, ki pobudo izvršijo z usklajenimi gibi govoril, kar zaznamo kot govor. Celotni proces govora je s tem opravljen, izvrši se le še preverjanje njegove pravilnosti. Ko nekaj izrečemo, izrečeno tudi slišimo, torej se izrečeno prenese skozi uho po slušnem živcu v slušno in Wernickovo središče ter v celotno možgansko skorjo, ki bodisi potrdi ali ovrže pravilnost izgovorjenega.

Za človeški govor je značilno, da je nadvse spremenljiv. Načini izgovaranja iste besede si niso nikoli povsem enaki. Spremenljivost govora v okviru istega besedila se lahko nanaša tako na različne govorce kot na enega govorca. Opisana spremenljivost govora je med drugim odvisna od neposredne komunikacijske

okolice, od hitrosti govora in od zapletenosti semantične vsebine besedila. Spremenljivost govora se izrazi v množici različnih govornih uresničitev načrtovanega besedila. Te se gibljejo v širokem razponu, od počasnega in nadvse jasnega govora do močno reduciranega in hitrega govora. Ta razpon je v strokovnem svetu poznan pod imenom hiper-hipo kontinuum govora [Lindblom 90].

1.2 Postopki sinteze govora

Z ozirom na zapletenost in nepopolno razumevanje pretvarjanja besedil v govor pri človeku se stroji za doseg istega cilja večinoma poslužujejo drugačnega načina. Kot pri človeškem branju si sledita dva procesa. Prvi, naravna obdelava jezika, besedilo prevede v ustrezni fonetični prepis, skupaj z zeleno intonacijo in ritmom. Drugi, digitalna obdelava signalov, preslika prejeto simbolično informacijo v govor.

Poglavitna razlika med kakšnim drugim govorečim strojem, npr. magnetofonom, ki predvaja vnaprej posnet signal, in sistemom, ki ga opisujemo, je v tem, da nas zanima samodejno tvorjenje *novih* stavkov. Sistemi, ki preprosto združujejo in predvajajo že vnaprej posnete besede oziroma dele stavkov, imenovani govorni odzivni sistemi [Sorin 94], so uporabni le, ko je seznam besed, ki ga je treba izgovoriti, omejen na manjše število, navadno na nekaj sto besed, oziroma kadar imajo stavki zelo omejeno skladiščno zgradbo, kot na primer pri samodejnem napovedovanju prihodov in odhodov vlakov na železniških postajah [Höge 94]. Pri samodejni pretvorbi neomejenih besedil v govor je nesmiselno posneti in hraniti vse besede danega jezika. Bolj primerno je, da samodejno pretvorbo besedil v govor definiramo kot samodejno tvorjenje govora, preko grafemsko-fonetičnega prepisa besedila, ki naj bo izgovarjano.

Prav tako je treba razlikovati med sistemi za pretvarjanje besedil v govor (angl. text-to-speech systems) in sistemi, ki v govor pretvarjajo vnaprej znane koncepte (angl. concept-to-speech systems) [Young 79; Pfister 94]. Prvim je na voljo le golo vhodno besedilo, medtem ko drugi dobijo na vhod semantične ali pragmatične koncepte skupaj s popolnim znanjem o pomenu stavka. Taki sistemi se uporabljajo, ko sistem sam sestavi besedilo, ki mora biti izgovorjeno, kot v primeru samodejnih sistemov za poizvedovanje po informacijah [Morton 93]. Opisani sintetizator govora nameravamo vključiti v modul za dialog samodejnega sistema za poizvedovanje o informacijah v letalskem prometu [Mihelič 94; Ipšič 97].

1.3 Sinteza govora za slovenski jezik

Prvi poskus v smeri sinteze slovenskega jezika je izvedel J. Hribar [Hribar 84].

Razvil je postopek za samodejno pretvarjanje slovenskega besedila v fonetični prepis besedila.

Razvoj sintetizatorja govora se je nadaljeval na Institutu Jožef Stefan v Ljubljani. S. Weilguny je, podobno kot pred njim Hribar, besedilo pretvoril v fonetični prepis, mu na zelo preprost način priredil prozodične parametre in ga posredoval dvema formantnima sintetizatorjema, razvitima za angleški jezik [Weilguny 93]. Raziskave na omenjenem inštitutu sta nadaljevala A. Dobnikar in T. Šef [Šef 96]. Razvili so difonski sintetizator govora [Šef 98], ki so ga pred nedavnim poklonili v uporabo slepim in slabovidnim osebam.

Tu velja omeniti tudi govoreči računalnik, ki ga je razvil in posebej prilagodil za potrebe slepih in slabovidnih A. Šurlan. Govorni del naprave je bil prvotno razvit za hrvaški jezik in je bil naknadno prilagojen slovenščini. Umetni govor je precej razumljiv, vendar je slišati preveč odrezavo in monotono, ker so vsi glasovi izgovorjeni na isti višini.

V Laboratoriju za umetno zaznavanje na Fakulteti za elektrotehniko v Ljubljani smo začeli razvijati sintetizator govora za slovenski jezik pred letom 1995. Kot cilj smo opredelili razvoj sintetizatorja govora, sposobnega kakovostnega samodejnega pretvarjanja slovenskih besedil v govor. Sistem naj bo sposoben prebrati poljubno besedilo in branje naj bo razumljivo ter karseda naravno. Zato je bilo treba poskrbeti za prijetno melodijo izgovora in se čim bolj približati pravilnemu naglaševanju besed. V nadaljevanju opisujemo, kako smo se lotili zastavljene naloge.

1.4 Zasnova sistema za sintezo slovenskega govora

Sistem za sintezo slovenskega govora smo zasnovali modularno. Tako lahko na preprost način zamenjamo en del sistema z novim, izboljšanim. Vhod v sistem predstavlja poljubno slovensko besedilo.

Prvi modul sistema opravi predobdelavo vhodnega besedila. V tem koraku se iz besedila odstranijo vsi odvečni simboli, ki ne vplivajo na izgovarjavo besedila, prav tako se števila razvijejo v ustrezne števnike. Nato se besedilo pretvori v ustrezen fonetični prepis. Če se vhodna beseda ne nahaja v slovarju izgovarjav, ji samodejno poiščemo naglasno mesto in jo pretvorimo v fonetični prepis.

Sledi nastavljanje vrednosti prozodičnih parametrov glasov. Sprva se določijo začetne, segmentne vrednosti trajanja in osnovne frekvence glasov. Nato se oblikuje besedni naglas in nazadnje stavčna intonacija.

Modul za združevanje osnovnih govornih enot izbere ustrezne difone, jih prilagodi želenemu trajanju in osnovni frekvenci ter jih združi v izhodni govorni signal.

Zamisel o modularni zgradbi sistema je bila prvotno realizirana v obliki večjega števila diplomskih nalog na Fakulteti za elektrotehniko v Ljubljani:

- grafemsko-fonetična pretvorba besedila T. Erpič [Erpič 95], R. Petrič [Petrič 97] in M. Burg [Burg 96], E. Tičević [Tičević 00];
- nastavljanje prozodičnih parametrov govora V. Vračar [Vračar 95], A. Rakar [Rakar 95], M. Benedik [Benedik 96] in A. Švigelj [Švigelj 97];
- izgradnja zbirke difonov in njihovo združevanje v govorni signal B. Grenc [Grenc 95], T. Šef [Šef 95], M. Sovdat [Sovdat 96] in T. Senčar [Senčar 96];
- izvedba preizkusa kakovosti sintetiziranega govora U. Mazej [Mazej 97].

V drugem poglavju opisujemo predobdelavo vhodnega besedila. V tretjem poglavju opisujemo postopek pretvarjanja predobdelanega grafemskega zapisa besedila v fonetični zapis.

Sledita dve poglavji, ki podajata opis prozodičnih lastnosti govornega signala in dognanj slovenskih jezikoslovcev, ki so obravnavali fonetiko in fonologijo slovenskega govora. Šesto poglavje opisuje meritve prozodičnih parametrov, ki smo jih izvedli na govornem signalu, da smo lahko postavili modele napovedovanja vrednosti prozodičnih parametrov za sintetični govor.

V sedmem poglavju opisujemo postopke napovedovanja vrednosti prozodičnih parametrov, ki jih nastavljamo sintetičnemu govoru. V naslednjem poglavju opisujemo način pridobivanja zbirke osnovnih govornih enot in podajamo opis postopka tvorjenja govornega signala z združevanjem govornih enot.

V devetem poglavju opisujemo potek preizkusa sintetizatorja govora, na katerem je 21 poslušalcev ocenjevalo naravnost in razumljivost sintetiziranega govora. Preizkus smo izvedli v skladu s priporočili mednarodnega združenja za telekomunikacije. Delo sklepamo s povzetkom opravljenega dela in z napotki za nadaljnje raziskave.

Uporabljene notacije

V nadaljevanju zapisujemo nize grafemov med lomljenimi oklepaji <>, nize fonemov med poševnimi oklepaji // in nize glasov med oglatimi oklepaji [], na primer: <kralj> niz grafemov, /kra:l/ niz fonemov, [kra:l'] niz glasov.

Oznake glasov zapisujemo v notaciji SAMPA.

2. PREDOBDELAVA VHODNEGA BESEDILA

Namen predobdelave ali normiranja vhodnega besedila je v tem, da se vhodno besedilo interpretira na enoličen način, in ga lahko izrazimo z zaporedjem besed v črkovni obliki in ločil. Ta del sistemov za sintezo govora je bil pogosto zanemarjan, v zadnjem času pa je deležen vedno večje pozornosti, predvsem v primerih uporabe sintetizatorja govora v komercialne namene [Gaved 93].

Mnogi sistemi za sintezo govora podpirajo aplikacije v telefoniji, kjer je pravilna izgovarjava uličnih in osebnih imen zelo pomembna, kot npr. v sistemu BT Laureate [Gaved 93]. Standardizirana izgovarjava krajevnih in lastnih imen ter imen ulic v evropskih jezikih je določena v okviru projekta ONOMASTICA [Schmidt 93].

V sistemu Infovox v različnih stopnjah predobdelave besedila dostopajo do mnogih virov znanja, predvsem sintaktičnega [Lindstrom 93]. V ta namen uporabljajo različne načine samodejne sintaktične razgradnje besedila. Ker je npr. pravilna izgovarjava števnikov pogosto odvisna od sobesedila, v sistemu SVOX z slovnico DCG (angl. definite clause grammar) razgradijo vhodno besedilo [Traber 93].

Vhod v sintetizator govora za slovenski jezik predstavlja poljubno slovensko besedilo, zapisano z ASCII znaki. Preden se besedilo pretvori v ustrezen fonetični prepis, ga je treba primerno predobdelati. Želimo, da je postopek predobdelave vhodnega besedila robusten in da ne predpostavlja strogega vnaprej določenega vhodnega formata besedila. Tako npr. dopuščamo možnost, da lastna imena niso vedno pisana z veliko začetnico ali pa da okrajšave niso vedno zaključene s piko. Kljub temu se mestoma vnaprej predpisanemu formatu ni bilo moč izogniti. V nadaljevanju opisujemo postopek predobdelave besedila. Posebni nečrkovni znaki imajo naslednji pomen: ~ (č), { (š), ` (ž), \ (tri pike),] (paragraf) in – (pomišljaj).

V postopku predobdelave vhodnega besedila iz vhodnega besedila odstranimo odvečne simbole, ki ne vplivajo na izgovarjavo besedila. Zaporedja števk razvijemo v primeren grafemski prepis (razvoj glavnih in vrstilnih števnikov), zaporedja velikih črk primerno obdelamo (ugotavljanje, ali gre za naslov ali za akronime). Za okrajšave določimo razširjeni grafemski zapis. Določimo način uporabe ločil (skladenjska ali neskladenjska raba...) in pretvorimo posebne simbole v opisni grafemski zapis.

Predobdelava vhodnega besedila poteka v dveh korakih: iskanje konca povedi in predobdelava besed v povedi.

Ker je besedilo, ki ga želimo sintetizirati, lahko poljubno dolgo, je smotno, da ga razbijemo na manjše enote, za katere določamo izgovarjavo. Za enoto smo

izbrali *poved* [Toporišič 91, str. 419]. Kot poved obravnavamo vsako zaporedje grafemov, ki se zaključijo z znakom za konec datoteke ali z enim izmed naštetih ločil, ki jih imenujemo *končna ločila*: pika v skladenjski rabi (.), klicaj (!), vprašaj (?), tri pike (...).

Določanje konca povedi oz. stavka ni tako preprosto, saj se ločila, ki lahko služijo kot končna ločila, ne uporabljajo vedno v skladenjski vlogi [Lieberman 89; SAZU 90, str. 37]. Za klicaj, vprašaj in tri pike v našem pristopu privzamemo, da gre za končno ločilo, medtem ko pika nima vedno vloge končnega ločila.

V tem podpoglavju opisujemo, na kakšen način pretvarjamo različne nize grafemov vhodnega besedila. Razlikujemo med naslednjimi nizi grafemov v vhodnem besedilu: niz malih črk, niz velikih črk, ločilo, niz števk, niz posebnih simbolov, kombinacije teh nizov. Vsakega izmed naštetih nizov grafemov obravnavamo na poseben način.

2.1 Črkovni grafemi

Nize malih črk, ki so ločeni z ločili, s presledki ali z znamenjem za prehod v novo vrsto, oblikujemo v besede, katerim kasneje priredimo ustrezno izgovarjavo. Ostale nize grafemov obravnavamo na naslednji način.

Kot besedo z velikimi črkami definiramo vsako besedo, ki vsebuje vsaj dve veliki črki. Besede z velikimi črkami uporabljamo [SAZU 90, str. 23]: pri poudarjanju besedil ali posameznih besed v njih (<PREPOVEDANO ODLAGANJE ODPADKOV>, <AVTOBUSNA POSTAJA>, <KUPIM nov žepni računalnik>), v naslovih, preglednicah ipd. (<MARTIN KR PAN>, <KONČNO OBLIKOVALI PREDLOG ZA NOVO VLADO>) in v kraticah ter kemijskih ali drugih simbolih (<DZ>, <HCOH>).

Besede z velikimi črkami obravnavamo na naslednji način. V primeru, da besedi z velikimi črkami sledi druga beseda z velikimi črkami, predpostavimo, da gre za naslov ali poudarjanje besedila, zato besedi pretvorimo v male črke: <DNEVNE NOVICE> v <dnevne novice>. Pred tem preverimo, ali se katera izmed besed nahaja v seznamu kratic ali akronimov.

V preostalih primerih gre za *kratico* ali *akronim*. Kratico pretvorimo na naslednji način. Najprej preverimo, ali se kratica nahaja v seznamu kratic. Ta seznam je razširljiv in vsebuje preko 900 različnih kratic, ki se pojavljajo pretežno v dnevnem tisku. Če je kratica daljša od treh črk, predpostavimo, da gre za nečrkovalno brano kratico, in jo izgovarjamo po enakih pravilih kot preostale besede: <UNICEF>, <NATO>. Sicer predpostavimo, da gre za črkovalno brano kratico, in jo črkujemo: <WHO>, <ZDA>.

2.2 Ločila

Ločila so nečrkovni grafemi, ki se uporabljajo na dva načina [SAZU 90, str. 37]: skladijsko (ločila zaznamujejo tonski potek, premore in vrste povedi ipd) in neskladijsko (ločila zaznamujejo okrajšave besed, vrstilstnost števnikov ipd.).

Skladijsko se uporabljajo: pika (.), klicaj (!), vprašaj (?), tri pike (...), vejica (,), podpičje (;), dvopičje (:), pomišljaj in vezaj (-), narekovaji ("', ') in oklepaji ().

Prva štiri ločila (!?...) v skladijski rabi zaznamujejo konec povedi, zato jih imenujemo *končna ločila*. Naslednjih pet ločil (,-;-) zaznamuje konec dela povedi.

Če želimo, da ločilo v besedilu pravilno upoštevamo, moramo sprva ugotoviti, ali je bilo ločilo uporabljeno skladijsko ali neskladijsko in v slednjem primeru določiti tudi natančen način uporabe. S tem znanjem lahko npr. ločujemo med piko, ki označuje konec povedi, in med decimalno piko v številki. Če se v vhodnem besedilu pojavijo narekovaji, jih pri izgovarjavi ne upoštevamo.

Vrsto rabe posameznih ločil določamo na naslednji način.

Piko obravnavamo kot končno ločilo v skladijski vlogi v primerih, ko:

- piki sledi presledek in beseda z veliko začetnico: <Dopolnil jih je 78. Lepa starost.>
- piki sledita dve znamenji za novo vrstico, kar sočasno zaznamuje konec paragrafa ,
- piki neposredno ali s presledkom ne sledi številka, saj gre v takem primeru za vrstilni številnik ali decimalno piko: <3.238>; <Ponovite 3. odstavek.>.

Pika ima poleg skladijske funkcije, kjer zaznamuje konec povedi, še tri neskladijske funkcije: označuje vrstilstnost števnika (Bil je 8. Marec.), decimalna pika (3.141592) in zaključek okrajšave (jaz oz. ti).

Zaključek okrajšave: če se pred piko nahaja niz, ki pripada seznamu okrajšav, se okrajšava pretvori v polno obliko, pika pa odstrani iz grafemskega niza.

Vrstilni številnik: če se pika pojavi neposredno za številko in ji ne sledi številka. V slednjem primeru je brez bolj natančnega poznavanja besedila težko razlikovati med decimalnim številom in zaporedjem vrstilnih števnikov. Odločili smo se za konvencijo, da mora v primeru vrstilnega števnika piki slediti presledek: <Bil je 5. januar.>, <Ob 8. uri zvečer.>. Če se pika pojavi za številko na koncu stavka, privzamemo, da gre za glavni številnik, in ne za vrstilnega: <Preostalo jih je le še 7.>.

Decimalna pika: če je pika v levem in desnem sosedstvu števk, privzamemo, da gre za decimalno piko, ki ločuje celoštevilčni del števila od decimalnega: <Število pi zapišemo na šest decimalnih mest natančno kot 3.141592.>.

Ločilo tri pike nastopi takrat, ko piki neposredno brez presledka sledita novi dve piki. V skladijski rabi so najbolj pogoste enodelne tri pike [SAZU 91, str. 50],

ki zaznamujejo nedokončano misel in tudi izpuščene dele navedkov (npr. <Saj bi mu povedal, pa...>), premor pri dodatnih pojasnitvah (npr. <Res..., iz navade!>) in spremembo skladijskega naklona v zloženi povedi.

V neskladenjski vlogi tri pike zaznamujejo izpust dela besede, npr. <O, pa saj to ni mogoč...>. Ločilo tri pike smo obravnavali kot končno ločilo v povedi.

Vprašaj in klicaj se večinoma uporabljata v skladijski vlogi in zaznamujeta konec povedi. Podpičje se uporablja samo skladijsko. Loči posamezne dele povedi bolj krepko kot vejica.

Vejica v skladijski rabi ločuje dele proste ali zložene povedi. Neskladenjsko rabo vejice ugotavljamo v številskih zvezah – decimalna vejica: če je vejica v levem in desnem sosedstvu števka, privzamemo, da gre za decimalno piko, ki ločuje celoštevilčni del števila od decimalnega: <Za dolar je treba odšteti že 1,71 nemške marke.>.

Ker vhod v sintetizator govora predstavlja zaporedje znakov ASCII, ne razlikujemo med pomišljajem in vezajem in vse znake '-' upoštevamo vedno kot pomišljaje. Pomišljaj v skladijski rabi: namesto vejice poudarjeno ločuje kakšno besedo ali misel stavka (<Ti očeta do praga – sin tebe čez prag!>), vpeljuje pojasnilo že zapisanega (<Le to je še pomembno – zdravje.>), zaznamuje nedorečeno misel, zamolk (<Saj bi jim povedal, pa –>).

Upoštevali smo dve neskladenjski rabi pomišljaja: matematični minus (če je desno od pomišljaja števka, privzamemo, da gre za matematični simbol minus: <3-5>, <-7>), znak za deljenje besede (če se pomišljaj pojavi na koncu vrstice, levostično naslonjen na predhodno besede, ga obravnavamo kot znak za deljenje ter združimo besedi na prelomu vrstice v eno besedo).

Čeprav se predklepaj in zaklepaj lahko pojavljata v skladijski in v neskladenjski rabi (kot del matematičnega izraza), ju obakrat obravnavamo na enak način.

2.3 Števke

Zaporedja števka ali cifer najpogosteje oblikujejo glavne oziroma vrstilne števnike.

Glavne števnike in decimalna števila pretvarjamo na naslednji način. Tisočice, stotice, desetice in enice se povežejo na ustrezen način. Glavne števnike od nič do sto pišemo skupaj, ostale narazen: <751312.123> se pretvori v <sedemsto enainpetdeset tisoč tristo dvanajst pika sto triindvajset>. Števnike nad milijon pretvorimo v zaporedje števka: <12345678> se pretvori v <ena dva tri štiri pet šest sedem osem>.

Podstavo *vrstilnega števnik*a dobimo na podoben način kot pri glavnem števniku. Ker je vrstilni števnik pridevniška beseda, je njegovo obrazilo odvisno od spola, sklona in števila besede, na katero se vrstilni števnik nanaša. Ker se odno-

snica lahko nahaja na drugem koncu povedi, je pravilna pretvorba vrstilnih števnikov nemogoča brez predhodne skladenjske razgradnje povedi.

Obrazilo vrstilnega števnik določimo na preprost način. Predpostavimo, da se vrstilni števnik nanaša na besedo, ki mu sledi. Zato se mora z njo ujemati v sklonu, spolu in številu. Opazujemo končaj te naslednje besede in sklepamo na obrazilo vrstilnega števnik po naslednjem vzorcu:

končaj besede za vrstilnim števnikom	obrazilo vrstilnega števnik	primer	
		ih	ih
ah	ih	na 4. dirkah	na četrtih dirkah
ev	ih	2. zvoncev	drugih zvoncev
ov	ih	otroci 8. razredov	otroci osmih razredov
ega	ega	3. slovenskega	tretjega slovenskega
esa	ega	1. kolesa	prvega kolesa
emu	emu	1. beneškemu	prvem beneškemu
em	im	pred 1. mesecem	pred prvim mesecem
im	im	s 5. moškim	s petim moškim
oma	ima	s 1. dečkoma	s prvima dečkoma
om	im	s 5. mestom	s petim mestom
ama	ima	s 1. nagradama	s prvima nagradama
ami	imi	s 1. nagradami	s prvimi nagradami
a	a	5. ulica	peta ulica
e	e	2. armade	druge armade
o	o	3. tekmo, 1. mesto	tretjo tekmo, prvo mesto
u	em	pri 1. tekaču	pri prvem tekaču
sicer	i	1. tolar	prvi tolar

Tabela 1: Primeri napovedovanja obrazila vrstilnega števnik na podlagi končaja naslednje besede.

Vsa obrazila vrstilnih števnikov se ne oblikujejo po takem vzorcu. Razloga za to sta dva. Beseda, ki sledi vrstilnemu števniku, ni njegova odnosnica. V takem primeru pomaga edino skladenjska analiza stavka, npr. <Nataša je bila le 3.>. Končaj besede, ki sledi vrstilnemu števniku, ne določa enoumno končnice vrstilnega števnik, npr.: <1. dete> in <1. ankete>. Obe besedi se končata na <e>, vendar sta vrstilna števnik različna: <prvo> in <prve>. V takem primeru je treba poznati spol, sklon in število odnosnice. Za to ponovno potrebujemo skladenjsko analizo ali pa potrebujemo slovar, v katerem se mora nahajati informacija o oblikoglasnih lastnostih besede.

Zato smo tabelo sestavili tako, da pokriva kar največ možnih kombinacij konča-

jev vrstilnih števnikov in njihovih odnosnic. V primeru ko enemu končaju odnosnice ustreza več možnih obrazil vrstilnega števnikar, smo se odločili za tisto, za katero smo menili, da nastopa bolj pogosto.

Pri pisanju datumov se za vrstilnim števnikom pogosto pojavi kateri od mesecev. V tem primeru opisana pravila ne veljajo povsem, zato smo dodali nova. Če vrstilnemu števniku sledi katerikoli izmed mesecev v poljubnem sklonu, sklepa-mo na obrazilo vrstilnega števnikar po naslednjem vzorcu:

končaj meseca za vrstilnim števnikom	obrazilo vrstilnega števnikar	primer	
r	i	3. oktober	tretji oktober
c	i	8. marec	osmi marec
l	i	7. april	sedmi april
j	i	2. julij	drugi julij
t	i	1. avgust	prvi avgust
a	ega	1. aprila	prvega aprila
u	em	po 2. marcu	po drugem marcu
em	im	s 5. januarjem	s petim januarjem
om	im	s 15. decembrom	s petnajstim decembrom

Tabela 2: Primeri napovedovanja obrazila vrstilnega števnikar na podlagi končaja mese-ca za vrstilnim števnikom.

Med grafemi poznamo še druga znamenja, predvsem ideograme, tj. znamenja, ki nosijo pomen, kot so npr. matematična znamenja, znamenja iz logike, izpostav-na znamenja, znamenja za vrednosti, mere, denarne enote, stopinje ipd.

V vhodnem besedilu upoštevamo še naslednje grafeme in jih pretvarjamo na naslednji način: & (in), ipd.

Kot primer podajamo rezultat predobdelave za naslednje vhodno besedilo:

PRED PRAZNIKOM

Pred soboto, 25. aprila, so imeli cvetličarji polne roke dela. Posebej živahno je bilo npr. v cvetličarnah GARDENIA in ROS. G. Pevec (podjetje GARDENIA) si je "mel roke", saj je izračunal, da bo izkupiček predprazničnega nakupa pre-cejšen, saj naj bi znašal 32000,00 SIT + 1000,00 SIT še neplačanih naročil – to pa je za 7 % več kot lani ...

Rezultat predobdelave vhodnega besedila; oznaka \ zaznamuje konec paragrafa, oznaka _ pa določa, da se mora beseda črkovati:

pred praznikom .]

pred soboto , petindvajsetega aprila , so imeli cvetli~arji polne roke dela .

posebej `ivahno je bilo naprimer v cvetli~arnah gardenia in R_O_S .
gospod pevec predklepaj podjetje gardenia zaklepaj si je mel roke , saj je
izra~unal, da bo izkupi~ek predprazni~nega nakupa precej{en, saj naj bi zna{al
dvaintrideset tiso~ vejica ni~ ni~ slovenskih tolarjev plus tiso~ vejica ni~ ni~
slovenskih tolarjev {e nepla~anih naro~il - to pa je za sedem procentov ve~ kot
lani \

Prizadevali smo si, da lahko obdelamo ~im bolj poljubno vhodno besedilo. V
primeru ko pa gre za konkretno aplikacijo, kjer sta podro~je uporabe in vhodni
formati vnaprej definirana, je mo~no predobdelavo besedila prilagoditi tako, da
~im bolj pravilno interpretira vhodno besedilo [Gaved 93].

3. PRETVARJANJE GRAFEMSKEGA ZAPISA BESED V FONETIČNEGA

Po končanem postopku predobdelave besedila je treba besedam prirediti primer-no izgovarjavo. Pri pretvarjanju grafemskega niza v fonetični prepis se uporabljajo naslednji pristopi: slovar izgovarjav kanoničnih oblik besed, uporaba pravil za grafemsko-fonetično pretvorbo, modeliranje pravil izgovarjave z binarnimi odločitvenimi drevesi, modeliranje izgovarjave besed z nevronskimi mrežami, modeliranje izgovarjave besed s stohastičnimi modeli, določanje izgovarjave z razgraditvijo besed na morfeme in uporaba pravil za fonetični prepis morfemov, določanje izgovarjave s pomočjo analogij in določanje izgovarjave jezikovno odvisnih besed z določanjem jezika govora.

Slovar izgovarjav nudi najbolj zanesljiv način za določanje fonetičnega prepisa kanoničnih besednih oblik [Dutoit 93, str. 101]. Žal slovar izgovarjav v mnogih primerih ni na voljo ali pa podaja izgovarjavo le za osnovno obliko besede in ne vsebuje izgovarjav preostalih pregibnih oblik besede.

Pravila za grafemsko-fonetično pretvorbo preslikajo grafemski niz v fonetični prepis [Bamakidis 87; Dutoit 93, str. 202]. Pridobimo jih lahko na več načinov: samodejno, z ekspertnimi sistemi, ali pa ročno [Bamakidis 87], na podlagi jezikoslovnega znanja in izkušenj. Ta pristop je uporaben predvsem za jezike, kjer je povezava med ortografskim in fonetičnim zapisom razmeroma preprosta, npr. za španščino ali finščino. Za večino ostalih jezikov velja, da mora slovar izgovarjav vsebovati vsaj tiste primere besed, katerih izgovarjave z uporabo pravil ni moč pravilno napovedati.

Ker je tvorba pravil izgovarjave naporna in zahteva poglobljeno jezikoslovno znanje, se v nekaterih sistemih uporabljajo podatkovno usmerjeni pristopi, ki temeljijo na nevronskih mrežah, stohastičnih modelih [Sagisaka 92] in teoriji o informacijah. Pravila izgovarjave pridobivajo tudi z binarnimi odločitvenimi drevesi, ki pri razraščanju kot kriterij uporabljajo medsebojno informacijo [Lucassen 84; Andersen 96].

Uporaba nevronskih mrež pri določanju izgovarjave besed, vsaj v primeru angleškega jezika [Ainsworth 89; McCulloch 87; Cohen 96; Adamson 96], ni prinesla pričakovanih rezultatov [Sproat 95, str. 204].

Pravila, ki določajo preslikavo morfemov, tj. najmanjših jezikovnih enot, ki še nosijo pomen, v fonetično obliko, se uporabljajo v sintetizatorjih govora za številne jezike. Prvič so bila uporabljena v sistemu MITalk [Allen 87] za angleški jezik, nato v sistemih PHRITTS [Meyer 93] in Syrub [Belhoula 94] za nemški jezik.

Razmeroma še neraziskan je pristop, ki se pri napovedovanju nepoznane izgovarjave besede naslanja na izgovarjavo njej podobne, a znane izgovarjave besede.

de [Dedina 91; Damper 96]. Če npr. slovar izgovarjav vsebuje izgovarjavo lastnega imena <Gorenc>, lahko sklepamo na izgovarjavo imena <Polenc> tako, da začetni glas [b] nadomestimo z glasom [p].

Drugačen pristop za določanje izgovarjave lastnih imen tujega izvora predvideva, da se s pomočjo jezikovnega modeliranja sprva ugotovi, kateremu jeziku beseda izvorno pripada, in se nato uporabijo pravila izgovarjave, značilna za tisti jezik [Gaved 93].

Pri razvoju sintetizatorja govora za slovenski jezik so se z izjemo [Šef 96] vedno uporabljala pravila za grafemsko-fonetično pretvorbo [Hribar 85; Weilguny 93; Erpič 95; Petrič 97].

Tudi v našem sintetizatorju govora uporabljamo pravila izgovarjave. Le-ta so skrbno izbrana in preverjena ter dopolnjevana na obsežnem slovarju najbolj pogostih izgovarjav za slovenski jezik. S statistično analizo slovarja izgovarjav smo uspeli izboljšati pravilnost določanja naglasnega mesta v besedi. Za razliko od ostalih pristopov za slovenski jezik določamo izgovarjavo besede v tekočem govoru in ne za kanonične oblike besed, izgovorjene v osami.

3.1 Določanje izgovarjave besed

Rezultat predobdelave vhodnega besedila je zaporedje besed in ločil, še vedno podanih v grafemskem zapisu. Sledi preslikava grafemskega zapisa besed v fonetični zapis. Pri tem si pomagamo s slovarjem izgovarjav besed in besednih zvez in s pravili, ki preslikajo grafemski niz v fonetični zapis. Besede se obdelujejo po vrsti od leve proti desni, ločila se prepisejo na izhod. Preslikava poteka na naslednji način:

1. iskanje besed in kolokacij v slovarju izgovarjav;
2. popravljanje izgovarjav besed in kolokacij z upoštevanjem koartikulacije med sosednjimi besedami;
3. določanje naglasnega mesta za besede, ki jih ni v slovarju izgovarjav, in
4. določanje fonetičnega zapisa besed, ki jih ni v slovarju izgovarjav, z upoštevanjem koartikulacije med sosednjimi besedami.

Besedo, ki je na vrsti za pretvorbo, poiščemo v abecedno urejenem seznamu slovarja najbolj pogostih izgovarjav. To pravzaprav ne velja le za eno samo besedo, temveč tudi za pogoste sklope besed, t. i. *kolokacije* [Smith 91, str. 76], kajti tudi izgovarjava poljubno dolgega sklopa besed se lahko nahaja v slovarju izgovarjav. Če sta beseda ali besedni sklop zastopana v slovarju, jima priredimo izgovarjavo iz slovarja.

Izgovarjavo besed ali kolokacij, ki se nahajajo v slovarju izgovarjav, je v nekaterih primerih treba uskladiti z izgovarjavo sosednjih besed. Za to uporabljamo

poseben seznam pravil za pretvorbo grafemov v fonetični zapis, ki deluje nad končaji besed in kolokacij.

Besedam, ko niso zastopane v slovarju izgovarjav, poskušamo določiti naglasno mesto. Sledi pretvarjanje grafemskega zapisa besede v fonetični zapis besede, pri čemer se upoštevajo pravila izgovarjave za posamezne besede in pravila, ki govori o vplivu koartikulacije pri izgovarjavi besede v sklopu s sosednjo besedo.

3.2 Iskanje besed in kolokacij v slovarju najbolj pogostih izgovarjav

Prvotna različica sintetizatorja govora za slovenski jezik je pri določanju izgovarjave besed uporabljala izključno pravila za pretvarjanje grafemskih nizov v fonetični zapis. Ker pa je naglasno mesto v slovenščini močno nepredvidljivo, je bila fonetična transkripcija besede pogosto nepravilna, kar je močno prispevalo k slabši razumljivosti sintetiziranega govora.

Vsaka izgovarjava besede, sorazmerno s svojo verjetnostjo nastopa besede, posredno vpliva na razumljivost in naravnost sintetiziranega govora. Zato je koristno, da imajo besede, ki jih shranimo v slovar izgovarjav, največje verjetnosti pojavljanja v slovenskem besedilu, njihova izgovarjava na izhodu sintetizatorja pa naj bo čim bolj razumljiva in naravna.

V nadaljevanju opisujemo, kako smo take besede poiskali s statistično obdelavo slovenskih besedil. Besedilo je v glavnem povzeto po diplomskem delu [Burg 96]. V zaprti domeni, za katero je značilno vnaprej določeno število besed ali celo fraz, se najpogosteje uporablja generiranje govora s snemanjem besed in fraz. Povprečna frekvenca besed je glede na ostale domene zelo visoka [Grabnar 96, str. 13]. Primer take uporabe je samodejno predvajanje opozoril v avtomobilih.

V polzaprti domeni sta vhodni format in vsebina dobro definirana, vendar obstaja možnost, da se kadarkoli pojavi nova beseda ali preneha uporabljati stara. Povprečna frekvenca nastopajočih besed je lahko zelo visoka. Taka primera sta govoreča vremenska napoved ali podajanje informacij o letalskem ali železniškem prometu.

Odperta domena je omejena le z besedami, ki jih pozna jezik. Besede imajo nizko povprečno frekvenco nastopanja.

Ker je sintetizator govora namenjen širšemu krogu ljudi, so besedila izbrana predvsem iz prispevkov v dnevnikih in tedenskih časopisih ter iz leposlovja. Statistično smo obdelali besedila, ki so vsebovala preko 800.000 besed. Pri tem smo ločeno obravnavali besede domačega izvora, besede tujega izvora (vsebujejo grafeme iz tujih abeced), kratice in lastna imena z veliko začetnico. Zbirka uporabljenih besedil je bila naslednja [Burg 96]:

	število besed
Sveto pismo	152.212
Mikeln, Veliki voz	162.396
Moj mikro, izbor člankov	150.194
Slovenec, izbor člankov	264.736
Jurčič, Deseti brat	65.860
Cankar, Moje življenje	26.916
skupaj	822.314

Tabela 3: Seznam besedil, ki smo jih pregleda pri iskanju najbolj pogostih besed.

Čeprav se je z omejitvijo področja, iz katerega besedila izhajajo (časopis, leposlovje), množica manj verjetnih besed zmanjšala, sodijo besede, s katerimi lahko generiramo celotno naše prebrano besedilo iz 822314 besed, v precej odprto domeno, saj je povprečna frekvenca pojavljanja besede 10.9.

Določeno število najbolj verjetnih besed se zaradi dovolj dolgega in dovolj splošnega besedila ne razlikuje od najbolj pogostih besed iz popolnoma odprte domene [Burg 96].

Z razmeroma majhnim številom najbolj pogostih besed, ki imajo širok razpon frekvenc, pokrijemo velik del prebranega besedila. Tako ima npr. 2100 najbolj pogostih besed, ki pokrivajo 68 % prebranega besedila, frekvenca nastopanja med 34 in 33.627. Število besed, potrebnih za pokritje novega odstotka prebranega besedila, se eksponentno povečuje [Burg 96, str. 18].

Med branjem vhodnega besedila se množica najbolj pogostih besed spreminja, prav tako se spreminja kumulativna verjetnost vseh besed v množici.

Krivulje, ki prikazujejo število najpogostejših besed in njihovo kumulativno verjetnost, so si za vse delne rezultate zelo podobne. Krivulja po prvih 111.000 prebranih besedah najbolj odstopa od ostalih, saj najpogostejših 3.000 besed pokriva 77 % prebranega besedila, najpogostejših 16.000 besed pa 93 %. To je 5 % več kot pri naslednjih delnih rezultatih pri 221.184, 373.857, 581.120, 665.901, 772.477 in 822.314 prebranih besedah. Slednji se med seboj zelo malo razlikujejo. Za vse nad 111.000 prebranih besed velja, da najpogostejših 3.000 besed pokriva 72 %, najpogostejših 16.000 pa 88.5 % besed celotnega besedila (Gros 97e, str. 19–22).

Tabela 4 podaja pokritost besedila s skupinami besed z določeno frekvenco. Frekvenčna območja so razdeljena binarno. Tretji stolpec podaja kumulativno verjetnost vseh besed znotraj binarnega področja, ta se giblje med 3 % in 7.6 %.

frekvenca besed	število besed	pokritost besedila [%]	pokritost besednjaka [%]
1	36660	4.6	48.620
2	11969	3.0	15.870
3 do 4	10114	4.3	13.410

5 do 8	7123	5.5	9.450
9 do 16	4451	6.5	5.900
17 do 32	2513	7.1	3.330
33 do 64	1361	7.6	1.810
65 do 128	632	7.1	0.840
129 do 256	299	6.8	0.400
257 do 512	145	6.2	0.190
513 do 1024	62	5.4	0.080
1025 do 2048	33	5.7	0.040
2049 do 4096	21	7.6	0.030
4097 do 8192	9	6.5	0.010
8193 do 16384	5	6.6	0.010
16385 do 32768	2	5.4	0.002
32769 do 65563	1	4.3	0.001

Tabela 4: Podatki za skupine besed s frekvenco v danem frekvenčnem območju. Drugi stolpec podaja število besed v skupini, tretji stolpec podaja pokritost besedila z vsemi besedami iz dane frekvenčne skupine (kumulativna verjetnost besed), četrty stolpec podaja delež besed iz dane frekvenčne skupine v besednjaku (kumulativna pokritost besednjaka).

Tabela 5 podaja najbolj pogoste besede v vhodnem besedilu skupaj z njihovo frekvenco. Opazimo, da so najbolj pogoste besede zelo kratke, saj so sestavljene le iz ene ali dveh črk. Med najbolj pogoste sodijo oblike pomožnega glagola biti (<je>, <so>), zaimki (<on>, <kaj>), vezniki (<in>, <pa>), predlogi (<v>, <na>, <za>) in členki (<ne>).

beseda	frekvenca	beseda	frekvenca	beseda	frekvenca
beseda	besede	beseda	besede	beseda	besede
je	33627	in	25040	v	18183
se	13679	da	11275	na	10120
so	9586	pa	8570	za	8063
ki	7592	ne	7513	z	5957
bi	5381	s	4898	tudi	4450
po	4351	ni	4286	ga	4060
to	3959	še	3761	bo	3549
ko	3517	sem	3416	ali	3414
bil	3054	tako	3037	iz	2993
od	2877	že	2491	mu	2482
pri	2471	če	2387	si	2342
kakor	2334	kaj	2294	bito	2254
vse	2212	jih	2106	o	2016

kot	2010	do	1977	kar	1922
lahko	1736	naj	1636	mi	1630
med	1622	več	1585	bila	1565
jo	1496	ker	1487	ti	1460
pred	1430	le	1418	sta	1394
ta	1390	tem	1365	k	1356
kako	1347	bodo	1335	smo	1268
nekaj	1230	ob	1227	zato	1191
jaz	1164	saj	1122	a	1096
me	1047	te	1039	samo	1037
tega	1028	rekel	1019	zdaj	1007
vam	1006	vendar	979	prav	957
jim	954	ter	920	bili	905
vas	898	svoje	852	namreč	808
sam	805	ste	767	potem	761
tam	760	vsi	760	bom	756
niso	752	pod	741	brez	740
veliko	735	zaradi	714	ima	706
nič	702	ampak	692	treba	691
gospod	691	kdo	681	nas	680
imel	667	dobro	664	zakaj	663
njim	645	dan	641	bolj	638
kjer	630	proti	615	tisti	613
bog	612	ji	612	res	596
vedno	593	nad	592	nisem	590
svojo	583	drugi	570	nam	566
spet	560	program	558	nato	557

Tabela 5: Seznam najbolj pogostih besed z njihovo frekvenco. Prebrana besedila so vsebovala 772.480 besed.

Prvotni približek izgovarjave besede smo določili samodejno z uporabo pravil za pretvarjanje grafemskega niza v fonetični zapis. Dobljene izgovarjave smo ročno pregledali in popravili.

Ista pisana beseda ima lahko več fonetičnih zapisov, ki se pogosto med seboj razlikujejo po sklonu in številu, lahko pa zapisi pripadajo tudi različnim besednim vrstam. Človek pri besedah, pri katerih so možne različne izgovarjave, iz sobesedila sklepa, kako pravilno naglasiti in izgovoriti besedo.

Za samodejno določanje pravilne izgovarjave bi bilo treba izvesti podrobno skladnjsko in semantično razgradnjo povedi, kdaj pa tudi pragmatično, saj ima npr. poved <On je.> glede na obe možni izgovarjavi [On jE] in [On je:] dva različna

pomena. Tega trenutna različica sintetizatorja govora ne omogoča, zato bodo napake, ki izvirajo iz tega naslova, ostale.

V primeru, ko ima beseda lahko več izgovarjav, smo se odločili za tisto, za katero smo domnevali, da nastopa bolj pogosto. Izgovarjave besed smo pisali v notaciji SAMPA. Slovar najbolj pogostih besed smo razdelili na več tematsko ločenih podslovarjev. Za določeno področje uporabe lahko dodamo nov podslovar ali dopolnimo katerega od obstoječih. Med določanjem izgovarjave program združi vse podslovarje v enoten slovar.

Slovarske besede so porazdeljene v sedmih podslovarjih, ki so opisani v tabeli 6. Trenutno je celotni slovar že dosti večji, saj obsega preko 40.000 najbolj pogostih slovenskih besed [Tičević 00].

	število besed
slovar kolokacij	17
slovar števil	234
slovar besed tujega izvora	304
slovar kratic	92
slovar lastnih imen	929
slovar ostalih besed	15.470
skupaj	16.215

Tabela 6: Seznam podslovarjev, ki tvorijo slovar pogostih izgovarjav.

Pravilna izgovarjava kolokacij lahko močno izboljša naravnost sintetiziranega govora [Lindstrom 93]. V človeškem govoru kolokacije nastopajo kot prozodične enote (najpogosteje ritmične enote) in so bolj podvržene redukciji in koartikulaciji znotraj kolokacije kot v primeru, da gre za ločene besede. Podobno kot Lindstrom v [Lindstrom 93] smo se odločili, da bomo izgovarjavo kolokacij zajeli v posebnem slovarju.

Nadalje lahko slovar kolokacij vsebuje zaporedja besed, katerih pomen je lahko dvoumen, če besede obravnavamo osamljeno, in katerih pomen postane določen, ko te besede nastopajo v kolokaciji. Tu se nahajajo tudi zaporedja besed, katerih izgovarjava se spremeni, če jih izgovorimo v kontekstu z drugo besedo, kot beseda <primer> in kolokacija <na primer>, kjer pride do premika naglasnega mesta. Nazadnje imamo naštetih nekaj primerov pridevnikov, ki v določenih besednih zvezah postanejo predslonke. Slovar kolokacij ali pogostih besednih zvez [Smith 91, str. 76] je še nepopoln, saj vsebuje le nekatere intuitivno izbrane besedne zveze [Gros 97e, str26].

V slovarju števnikov se nahaja izgovarjava pogostih glavnih, vrstilnih in množilnih števnikov. Glavni in vrstilni števniki so večinoma zajeti v razponu od 0 do 100, naprej pa v stoticah. Slovar kratic obsega nekatere bolj pogoste kratice, ki se pojavljajo predvsem v dnevnem in strokovnem tisku. V slovarju lastnih imen

se nahaja izgovarjava pogostejših osebnih in krajevnih lastnih imen. Slovar besed tujega porekla obsega besede, ki v grafemskem zapisu vsebujejo znake tuje abecede. Dodane so tudi nekatere druge besede tujega izvora ter nekateri izrazi iz računalniške stroke. Slovar preostalih besed je najbolj obsežen in vsebuje izgovarjave ostalih najbolj pogostih besed slovenskega izvora, na primer:

	izgovarjava
glasnim	gla:snim
glasno	gla:snO
glasom	gla:sOm
glasov	glasO:w
glasovi	glasO:vi
glasu	glasu:
glav	gla:w

Tabela 7: Primer zapisa iz slovarja pogostih izgovarjav.

Kanonična oblika izgovarjave podaja tisto izgovarjavo besede, ki jo izgovorimo v osami [Wesenick 96]. Ko isto besedo izgovorimo v zveznem govoru, v kontekstu, njena izgovarjava pogosto odstopa od kanonične oblike izgovarjave besede. Ta odstopanja lahko zajamemo s pravili, ki jih naknadno uporabimo nad kanonično obliko izgovarjave besed.

Če smo izgovarjavo besede ali kolokacije pridobili iz slovarja izgovarjav, jo moramo ustrezno popraviti. Pravila uporabimo nad kanoničnimi oblikami besed na način, kot je to opisano v nadaljevanju poglavja (tam najdemo tudi razlago posebnih oznak v spodnji tabeli). V tem poglavju je opisan tudi pomen posebnih simbolov, ki nastopajo v pravilih. Pravila, ki smo jih upoštevali, zajemajo pojave, ki se pri glasovih pojavljajo zaradi soseščine drugih glasov, in jih obravnava t. i. sintagmatika fonemov [Toporišič 91, str. 95]:

končaj besede	desni kontekst	fonetični prepis
v	—	[W]
v	—	[w]
b	—	[p]
d	—	[t]
g	—	[k]
z	—	[s]
`	—	[S]
p	—*	[b]
f	—*	[v]
t	—*	[d]
s	—*	[z]

{	_*	[ʔ]
c	_*	[dz]
~	_*	[dZ]
k	_*	[g]
m	_f	[F]
m	_v	[F]
n	_g	[N]
n	_k	[N]
n	_h	[N]

Tabela 8: Primeri fonetičnega prepisa končajev besed.

Predvsem smo upoštevali pravilo, ki predpisuje, da lahko pred zvenečim nezvočnikom stoji le zveneči nezvočnik in pred nezvenečim spet le nezveneči. Nadalje se <n> pred <k, g, h> izgovarja mehkonebno; <m> pred <f, v> pa zobnoustnično.

3.3 Pretvarjanje grafemskega zapisa besed v fonetični prepis

Pretvorbo besedila v zaporedje glasov lahko opišemo kot zaporedje dveh preslikav – določanja naglasnega mesta in določanja izgovarjave besed. Pri zapisu fonetičnega prepisa besed smo odločili, da bomo iz praktičnih razlogov namesto fonetičnih simbolov IPA (International Phonetic Alphabet) uporabili računalniško berljivi seznam fonetičnih simbolov SAMPA (SAM (Speech Assessment Methods) Phonetic Alphabet) [Wells 89]. Seznam simbolov SAMPA za slovenski jezik je še v pripravi [Dobrišek 96b].

ime glasu	SAMPA simbol	primer
a	[a]	kazen
b	[b]	bog
c	[ts]	cena
č	[tS]	čovelj
d	[d]	dar
široki e	[E]	cesar
ozki e	[e]	cena
f	[f]	fant
g	[g]	gora
h	[h]	hiša
i	[i]	ime
j	[j]	jok
k	[k]	kol

l	[l]	lice
mehčani l	[l']	valj
m	[m]	miza
zobnoustnični m	[F]	nimfa
n	[n]	nos
mehčani n	[n']	vonj
mehkonebni n	[N]	tank
široki o	[O]	koš
ozki o	[o]	koliko
p	[p]	past
r	[r]	res
s	[s]	senca
š	[S]	šilo
t	[t]	tvoj
u	[u]	ulica
v	[v]	vaza
dvoglasniški in zvoneči dvoustnični		
nedvoglasniški	[w]	dal, vzev, vstav
nezvoneči dvoustnični		
nedvoglasniški	[W]	vsak, predvsem, ustrašiti
z	[z]	zven
ž	[Z]	žamet

Tabela 9: Seznam SAMPA fonetičnih simbolov za slovenski jezik.

Posebni nečrkovni znaki imajo naslednji pomen: " kračina, : dolžina, | dvopičje, / tri pike, – pomišljaj in \$ paragraf.

3.3.1 Napovedovanje naglasnega mesta

Naglas je značilnost besede, medtem ko je poudarek značilnost stavka [Toporišič 91, str. 57]. Glede na število naglasov se besede delijo na: besede brez naglasa: breznaglasnice (enklitike in proklitike), besede z enim naglasom in besede z več naglasi. Dva ali več naglasov imajo le tisti sklopi, pri katerih se pomena posameznih sestavnih delov dobro zavedamo, npr. <ne'bodiga'treba>. Dva naglasa ima skupina zloženk in sestavljenk, več kot dva pa samo priredne zloženske.

Večzložna beseda, ki ni naglašena na zadnjem zlogu, se imenuje *bariton*. Besede

da, ki je naglašena na zadnjem zlogu, je *oksiton*, lahko je enozložna ali večzložna.

Mesto naglasa določa zlog, na katerem ima naglašena beseda jakostno ali tonemsko izrazitost. S tem vpliva tudi na vse preostale zloge. V nekaterih jezikih ima naglas stalno mesto, npr. v francoščini je na zadnjem zlogu, v poljščini na predzadnjem zlogu, v češčini na prvem zlogu..., ali pa delno omejeno mesto naglasa; primer je hrvaščina, kjer ni nikoli naglasa na zadnjem zlogu, in prosto mesto naglasa, npr. v slovenščini.

V grščini je naglasno mesto podano že v samem grafemskem zapisu besede. V slovenščini pa se lahko naglas nahaja na kateremkoli zlogu v besedi [Omerza 70, str. 116]. Naglas je možen na prvem zlogu, zadnjem zlogu, predzadnjem zlogu in na predpredzadnjem zlogu.

Če je beseda dvozložna, tedaj je en zlog naglašen in drugi nenaglašen ali nasprotno. Čeprav je v slovenščini naglas pogosto na prvem zlogu, kot npr. v besedah <vaja>, <pesem>, <vino>, poznamo tudi mnogo dvozložnih besed, ki imajo naglas na drugem zlogu, v tem primeru na zadnjem zlogu, npr. besede <drevo>, <lepo>, <sani>.

Če je beseda trizložna, je najbolj pogost naglas na drugem zlogu, npr. v besedah <velika>, <beseda>, <prometen>. Včasih je naglašen prvi zlog, zlasti v izpeljankah, npr. v besedah <čajnica>, <mišnica>, <okence>. Malo pa je trozložnih slovenskih besed domačega izvora, ki bi bile naglašene na tretjem ali v tem primeru na zadnjem zlogu, kot so npr. <plapola>, <drgeta>, <šelesti>. Podobno velja za besede, ki imajo več kot tri zloge.

Naglasno mesto je v slovenščini le deloma predvidljivo. Določajo ga naslednja pravila [SAZU 90, str. 108]:

- Pravila o vedno naglašanih končnicah, obrazilih in končajih: slovenski prapovpis [SAZU 90, str. 108] in slovnica slovenskega jezika [Toporišič 91, str. 53–56] podajata sezname delov besed, ki so skoraj vedno naglašeni. Vendar se pravila nanašajo na primere, ko je besedna oblika točno določena, npr. med stalno naglašene je uvrščena končnica <-a> v rodilniku množine samostalnikov ženskega spola <žen'a>.
- Pravila o nenaglasljivih enotah: nekateri končaji besed so le redko naglašeni.
- Pravila o naglasnih tipih: glede na lego naglasa v različnih oblikah besede razlikujemo štiri osnovne naglasne tipe: nepremični, premični, končniški in mešani naglas. S poznavanjem besedne vrste, pravil pregibanja in naglasnega tipa lahko za besede, podane v njihovi osnovni obliki (npr. imenovalnik in rodilnik pri samostalnikih, nedoločnik in sedanjik pri glagolu ipd.), sklepamo na mesto naglasa besede v vseh njenih izpeljanih oblikah.

Za uporabo slednjega pristopa je treba imeti na voljo slovar vseh slovenskih besed v njihovih osnovnih oblikah, skupaj z informacijo o načinu pregibanja in naglasnem tipu. Nad besedno obliko, ki ji želimo določiti naglasno mesto, se

sprva izvede postopek oblikoslovne analize, ki se imenuje geslenje ali lematizacija. Za dano besedno obliko se s pomočjo paradigem poišče v slovarju ustrezna lema. Sledi določanje naglasnega mesta z uporabo naglasnih shem. Samodejni postopek lematizacije za slovenski jezik so razvili na Inštitutu Jožef Stefan [Erjavec 92; Erjavec 97]. Uporabili so ga za določanje naglasnega mesta pri grafičnem fonemski pretvorbi [Weilguny 92]. Žal je postopek razmeroma počasen in trenutno še ne pokriva vseh besednih vrst. Prav tako nimamo na voljo slovarja osnovnih besednih oblik z označenimi oblikoglasnimi lastnostmi.

Pri ugotavljanju naglasnega mesta smo po svoji presoji uporabili seznama pravil o vedno naglašanih in nenaglasljivih enotah, ki smo jih sestavili iz različnih virov [Toporišič 91, str. 53–56; SAZU 90, str. 108]. Dodali smo tudi nekaj svojih pravil. Do končne zbirke pravil smo prišli z večkratnim preverjanjem pravilnosti določanja naglasnega mesta na obsežnem seznamu besed z več kot 16.000 različnimi najbolj pogostimi slovenskimi besedami.

Sestavili smo pet seznamov besed: seznam pripon, ki navadno niso nikoli naglašene, seznam predpon, ki večinoma niso naglašene, seznam začetnic, ki so večinoma naglašene na prvem zlogu, seznam končajeve besed s pogostim naglasom na predpredzadnjem zlogu, seznam končajeve besed s pogostim naglasom na predzadnjem zlogu, seznam končajeve besed s pogostim naglasom na zadnjem zlogu in seznama enklitik in proklitik, ki niso naglašene. Sezname so podrobno opisani v navedeni literaturi [Burg 96] in [Gros 97e, str. 33–35].

Postopek določanja naglasnega mesta je podan v navedeni literaturi [Gros 97e, str. 37].

Ker v postopku napovedovanja naglasnega mesta ne uspe vedno najti naglasnega mesta, smo morali poiskati pravila za samodejno naglaševanje preostalih besed. Z analizo slovarja izgovorjav 16.000 najbolj pogostih besed smo ugotovili, da je skoraj 6.000 najbolj pogostih besed trozložnih. Po pogostosti nastopanja jim sledijo dvozložne in štirizložne besede.

število zlogov v besedi	število besed z naglasom na					
	1. zlogu	2. zlogu	3. zlogu	4. zlogu	5. zlogu	skupaj
1	662					662
2	3568	899				4467
3	1178	4548	237			5963
4	72	1301	1347	49		2769
5 in več	5	54	282	290	3	634

Tabela 10: Odstotek naglasov na določenem zlogu v besedah z različnim številom zlogov.

Z dobljenimi rezultati smo napovedali najbolj verjetno naglasno mesto za posamezno besedo glede na število zlogov. Če je beseda dvozložna, je najbolj verjet-

no, da je naglašena na prvem zlogu. Trozložne besede so najbolj verjetno naglašene na drugem zlogu, štirizložne besede na drugem ali tretjem zlogu, pet- ter večzložne besede pa na tretjem ali četrtem zlogu.

Pravilnost napovedovanja naglasnega mesta lahko še izboljšamo, če pri statistični analizi upoštevamo tudi strukture zlogov [Tičević 00].

3.3.2 Določanje fonetičnega zapisa besed

Po določanju naglasnega mesta za besede, ki niso zastopane v slovarju izgovarjav, sledi samodejno pretvarjanje grafemskega zapisa besede v fonetični zapis besede. Pri tem se upoštevajo pravila izgovarjave za posamezne besede in pravila, ki govorijo o vplivu koartikulacije pri izgovarjavi besede v sklopu s sosednjo besedo.

Za pretvarjanje grafemskih nizov v fonetični prepis uporabljamo dve vrsti pravil: kontekstno neodvisna pravila in kontekstno odvisna pravila, kjer je preslika grafskega niza odvisna od njegovega levega in desnega konteksta.

Kontekstno neodvisna pravila enolično preslikajo grafemski niz v fonetični simbol. Večina pravil za grafemsko-fonetično pretvorbo danega niza v slovenskem jeziku je kontekstno odvisnih. To pomeni, da se grafem oziroma grafemski niz lahko preslika v različne fonetične simbole, odvisno od konteksta, v katerem se nahaja. Pravila, ki določajo, kateri alofon določenega fonema velja uporabiti v določenem fonetičnem prepisu, so tudi kontekstno odvisna. To je razumljivo, saj je pojavljanje alofonskih različic fonema odvisno od glasovnega okolja, v katerem se alofoni nahajajo.

Kontekstno odvisno pravilo sestavljajo niz grafemov, ki ga preslikujemo, njegov levi in desni kontekst ter niz fonetičnih simbolov, v katerega se preslika grafemski niz, na primer:

levi kontekst	niz grafemov	desni kontekst	fonetični prepis	primer
\$	ev	–	[@w]	čevljev
\$	er	–	[@r]	gaber

Tabela 11: Primer pravila za pretvarjanje grafemskega niza v fonetični prepis.

Izrazno moč pravil še povečamo, če pri vhodnem seznamu grafemskih znakov uvedemo dodatne oznake, s katerimi ponazorimo pravila, ki delujejo nad večjo skupino glasov. Vsaka dodatna oznaka opisuje določeno podmnožico vhodne abecede. Dodatne oznake dopuščamo le za opisovanje konteksta, ne pa za opisovanje sredine pravila. Oznaka '/' zaznamuje naglasno mesto, oznaka '_' pa presledek:

oznaka za glas. skupino	glasovna skupina	glasovi iz glasovne skupine
#	samoglasniki	aeiou
\$	soglasniki	mnvljrbdz'gpftcs~{kh
&	soglasniki, skupaj s presledkom	mnvljrbdz'gpftcs~{k h
%	zvočniki	mnvjlr
"	nezveneči soglasniki	pftcs~{kh
+	zveneči soglasniki	bdz'gmnvjlr
*	zveneči nezvočniki	bdgz'
=	vsi glasovi, skupaj z '/' in ' _ '	abc~defghijklmnoprs{tuvz`xywq/_

Tabela 12: Opis posebnih znakov, ki jih uporabljamo v pravilih za grafemsko-fonetični prepis.

Grafemski zapis vhodnega besedila se pretvarja od leve proti desni in se ne vrača nazaj. Pravila se pregledujejo po vrsti, zato si morajo slediti tako, da so na začetku seznama pravil za posamezen grafem tista, ki opisujejo posebne primere, sledijo pa jim bolj splošna pravila. Izgovarjave besed, ki jih ne moremo opisati z izbrano zbirko pravil, posebej navedemo v slovarju izgovarjav. Zaporedje pravil je naslednje: pravila za polglasnik, pravila za samoglasnike in pravila za soglasnike.

Pri sestavljanju zbirke pravil smo izhajali iz predlogov, podanih pri Hribarju [Hribar 85] in Erpiču [Erpič 95]. Ti predlogi se zgledujejo po pravilih, podanih v slovenski slovnici [Toporišič 91]. Kontekstno neodvisna pravila smo spremenili v kontekstno odvisna, tako da smo v pravilu levemu in desnemu kontekstu privedili poljuben kontekst (znak =). S pregledovanjem učinka pravil na bazi 16.000 izgovarjav različnih slovenskih besed in nenehnim dopolnjevanjem in popravljanjem pravil smo prišli do končne zbirke 143 kontekstno odvisnih pravil.

levi kontekst	niz grafemov	desni kontekst	fonetični prepis	primer	razlaga pravila
\$	ev	_	[@w]	žebļjev	polglasnik se pridruži vsakemu -r, ki mu ne sledi samoglasnik [Toporišič 91, str. 49]
\$	er	_	[@r]	gaber	
\$	el	_	[@w]	padel	polglasnik se pojavlja v predsamoglasniškem sklopu nezvočnik + e + (-lj ali -nj), če tak sklop preide v položaj na koncu besede ali pred soglasnik iste besede [Toporišič 91, str. 49]
\$	elj	_	[@l']	čevelj	

\$	enj	_	[@n']	semenj	ista razlaga kot zgoraj
r	en	&	[@n]	kurent	
\$	ec	_	[@c]	starec	polglasnik se pojavlja pred samostalniškima priponama -c in -k ali pred pridevniškima priponama -k in -n v imenovalniku ednine [Toporišič 91, str. 49]
\$	en	_	[@n]	česen	ista razlaga kot zgoraj
\$	ek	_	[@k]	dedek	ista razlaga kot zgoraj
\$	em	_	[@m]	sedem	
/#%	e	%	[@]	sejem	polglasnik se vriva v večini sklopov zvočnik + e + zvočnik, kakor hitro pride v položaj na koncu besede ali pred soglasnikom [Toporišič 91, str. 50]
/#\$	e	#\$	[@]	pisemski	polglasnik se pojavlja v predsamoglasniškem sklopu nezvočnik + e + zvočnik, če tak sklop preide v položaj na koncu besede ali pred soglasnik iste besede [Toporišič 91, str. 50]
=	/e	nt#	[E:]	študenta	širok je naglašeni <e> v priponi -ent- pred samoglasnikom in na koncu besede (izjeme trobenta, polenta) [Toporišič 91, str. 46]
=	/e	nt_	[E:]	študent	ista razlaga kot zgoraj
=	/e	j#	[E:]	veja	
_t	/e	b#	[E:]	tebe	širok je naglašeni <e> v oziralnih zaimkih za prvo in drugo osebo ednine
_m	/e	n#	[E:]	mene	ista razlaga kot zgoraj
_	/e	n#	[E:]	ena	širok je naglašeni <e> v sklonih glavnega števnika en
l	/e	_	[E"]	kajle	
=	/e	=	[e:]	pena	preostali neglašeni <e> so ozki
=	e	=	[E]	drevo	nenaglašeni <e> so široki
=	/o	ba	[O:]	svetloba	pripone -oba, -oča, -ota pri samostalniku in -ok, -oka pri pridevniku, imajo široki [O:] [Toporišič 91, str. 47]
=	/o	~a	[O:]	čistoča	ista razlaga kot zgoraj
=	/o	ta	[O:]	trdota	ista razlaga kot zgoraj

=	/o	ka	[O:]	široka	ista razlaga kot zgoraj
=	/o	kega	[O:]	globo- -kega	ista razlaga kot zgoraj
=	/o	v#	[O:]	njegova, sova	v moderni izgovarjavi je brez izjeme širok vsak <o> pred <v>, ki mu sledi samoglasnik [Toporišič 91, str. 47]
_m	/o	j#	[O:]	moji	širok je naglašeni <o> v svojilnih zaimkih za prvo in drugo osebo ednine
_tv	/o	j#	[O:]	tvoje	ista razlaga kot zgoraj
=	/o	=	[o:]	pilot	preostali naglašeni <o> ji so ozki
=	o	=	[O]	korak	nenaglašeni <o> ji so široki
=	i	#	[ij]	radio	v izgovarjavo sklopa <i + samoglasnik> se navadno vrine [j] [SAZU 90, str. 84]
=	/i	=	[i:]	pismo	
=	i	=	[i]	primer	
_n	/a	j	[a"]	največ	kratki naglašeni
=	/a	=	[a:]	kazen	dolgi naglašeni
=	a	=	[a]	kocka	nenaglašeni
_	u	"	[W]	vsak	<u> kot predpona se pred nezvenečim nezvočnikom izgovarja kot nezveneči dvoustnični nedvoglasniški [W] [SAZU 90, str. 147]
_	u	+	[w]	ubiti	<u> kot predpona se izgovarja kot zveneči dvoustnični nedvoglasniški [w] [SAZU 90, str. 147]
#	u	#	[w]	Dachaua	<u> v polcitatnih besedah se izgovarja kot dvoglasniški [w] [SAZU 90, str. 147]
=	/u	=	[u:]	ulica	naglašeni <u>
=	u	=	[u]	kurir	nenaglašeni <u>
=	m	v	[F]	žemva	<m> pred <v> in <f> se izgovarja zobnoustnično [SAZU 90, str. 145]
=	m	f	[F]	simfonija	velja ista razlaga kot zgoraj
=	m	_v	[F]	sem videl	velja ista razlaga kot zgoraj
=	m	_f	[F]	sem fant	velja ista razlaga kot zgoraj
=	mm	=	[m]	gamma	v besedah tujega izvora
=	m	=	[m]	marec	
=	nj	&	[n']	vonj,	

				konjski	sklop <nj> pred soglasnikom ali na koncu besede se izgovarja podaljšano ali neobvezno mehčano [SAZU 90, str. 141]
\$	nj	#	[nj]	marnja	
=	n	k	[N]	Anka	<n> pred <k>, <g> ali <h> se izgovarja mehkonebno [SAZU 90, str. 141]
=	n	g	[N]	angel	velja ista razlaga kot zgoraj
=	n	h	[N]	Anhovo	velja ista razlaga kot zgoraj
=	n	_k	[N]	en kos	velja ista razlaga kot zgoraj
=	n	_g	[N]	on gre	velja ista razlaga kot zgoraj
=	n	_h	[N]	en hip	velja ista razlaga kot zgoraj
\$	n	&	[@n]	tovarn, marnj	v mlajših prevzetih besedah lahko pride do neobveznega polglasnika med zvočnikoma (<l>, <n>, <v> ali <j>) na prvem mestu in zvočnikom (<m>, <n>, <r> ali <l>) na drugem mestu [Toporišič 91, str. 51]
=	nn	=	[n]	vanno	besede tujega izvora
=	n	=	[n]	noč	
=	lj	&	[l']	valj, boljši	sklop <lj> pred soglasnikom ali na koncu besede se izgovarja podaljšano ali neobvezno mehčano [SAZU 90, str. 141]
\$	lj	#	[lj]	žemlja	
\$	l	&	[@l]	Karlsbad	v mlajših prevzetih besedah lahko pride do neobveznega polglasnika med zvočnikoma (<l>, <n>, <v> ali <j>) na prvem mestu in zvočnikom (<m>, <n>, <r> ali <l>) na drugem mestu [Toporišič 91, str. 51]
=	l	_	[w]	tekel	na koncu besede <l> izgovarjamo kot dvoglasniški [w] [SAZU 90, str. 147]
=	ll	=	[l]	Minelli	v besedah tujega izvora
=	l	=	[l]	leto	
&	/r	&	[@"r]	rja, črna	<r> na začetku besede se pretvori v kratki [naglašeni polglasnik + r] v primeru, da je naglašeni zlog ne vsebuje nobenega samoglasnika

&	r	&	[@r]	rdeč, koprni, žanr	polglasnik se obvezno pojavi ob <r>, če le-temu ne sledi samoglasnik [Toporišič 91, str. 49]
=	/r	=	[r]		
=	rr	=	[r]	Cerro Torre	v besedah tujega izvora
=	r	=	[r]	škorec	
=	v	"	[W]	v petek	
=	v	"	[w]	siv	v dvoglasniškem sklopu in na koncu besede se <v> izgovarja dvoglasniško [SAZU 90, str. 147]
=	v	"	[W]	vsak	v nedvoglasniškem sklopu se <v> pred nezvenečim nezvočnikom izgovarja kot nezveneči dvoustnični nedvoglasniški [W] [SAZU 90, str. 147]
=	v	+	[w]	vzeti	v nedvoglasniškem sklopu se <v> izgovarja kot zveneči dvoustnični nedvoglasniški [w] [SAZU 90, str. 147]
=	v	=	[v]	vaza	navadni zobnoustnični [v]
=	w	"	[W]		v besedah tujega izvora
=	w	=	[w]		v besedah tujega izvora
=	s~	=	[StS]	sčasoma	v sklopu sičnik šumevec se navadno sičnik izgovarja kot ustrezni šumevec [SAZU 90, str. 84]
=	sch	=	[S]	Schultz	v besedah tujega izvora
=	tsch	=	[tS]	Deutsche Telekom	v besedah tujega izvora
=	ch	=	[k]	Christina	v besedah tujega izvora
=	kh	=	[k]	Akhbar	v besedah tujega izvora
=	th	=	[t]	Thebe	v besedah tujega izvora
=	ph	=	[f]	Philip	v besedah tujega izvora
=	cc	=	[k]	Mecca	v besedah tujega izvora
=	b	"	[p]	obtok	če zvenečemu nezvočniku sledi nezveneči nezvočnik ali premor, pride do premene v zvenečnosti [SAZU 90, str. 83]
=	d	"	[t]	odtok	velja ista razlaga kot zgoraj
=	g	"	[k]	krogca	velja ista razlaga kot zgoraj

=	z	"	[s]	razseliti	velja ista razlaga kot zgoraj
=	`	"	[S]	ježka	velja ista razlaga kot zgoraj
=	b	_"	[p]	ob toku	velja ista razlaga kot zgoraj
=	d	_"	[t]	od tega	velja ista razlaga kot zgoraj
=	g	_"	[k]	sneg topi	velja ista razlaga kot zgoraj
=	z	_"	[s]	mraz terja	velja ista razlaga kot zgoraj
=	{	_"	[S]	mož teče	velja ista razlaga kot zgoraj
=	p	*	[b]	Campbell	če nezvenečemu nezvočniku sledi zveneči zveneči nezvočnik ali pre mor, pride do premene v zvenečnosti [SAZU 90, str. 83]
=	f	*	[v]	Afgan	velja ista razlaga kot zgoraj
=	t	*	[d]	svatba	velja ista razlaga kot zgoraj
=	k	*	[g]	kdaj	velja ista razlaga kot zgoraj
=	s	*	[z]	glasba	velja ista razlaga kot zgoraj
=	{	*	[S]	vešda	velja ista razlaga kot zgoraj
=	c	*	[dz]	Kocbek	velja ista razlaga kot zgoraj
=	~	*	[dZ]	odločba	velja ista razlaga kot zgoraj
=	p	_*	[b]	rop banke	velja ista razlaga kot zgoraj
=	f	_*	[v]	grof gre	velja ista razlaga kot zgoraj
=	t	_*	[d]	kot gre	velja ista razlaga kot zgoraj
=	k	_*	[g]	vsak dan	velja ista razlaga kot zgoraj
=	s	_*	[z]	kos žemlje	velja ista razlaga kot zgoraj
=	{	_*	[Z]	koš gnoja	velja ista razlaga kot zgoraj
=	c	_*	[dz]	stric gre	velja ista razlaga kot zgoraj
=	~	_*	[dZ]	fantič gre	velja ista razlaga kot zgoraj
=	b	=	[b]	brez	
=	c	=	[ts]	konec	
=	~	=	[tS]	vrač	
=	d	=	[d]	denar	
=	ff	=	[f]		
=	f	=	[f]	fant	
=	gg	=	[g]	loggia	
=	g	=	[g]	gams	
=	h	=	[h]	hitro	
=	j	=	[j]	jesen	

=	kk	=	[k]	mokka
=	k	=	[k]	pek
=	p	=	[p]	pest
=	qu	=	[k]	Monique
=	ss	=	[s]	Rossi
=	s	=	[s]	kosa
=	{	=	[S]	šala
=	tt	=	[t]	Pavarotti
=	t	=	[t]	telo
=	x	=	[ks]	Obelix
=	y	=	[i]	Orly
=	zz	=	[ts]	palazzo
=	z	=	[z]	zima
=	`	=	[Z]	življenje

Tabela 13: Pravila za pretvarjanje grafemskih nizov v fonetični prepis.

Črkovalno brane kratice lahko naglašujemo na vsakem zlogu posebej ali pa le na končnem zlogu [SAZU 90, str. 125]. V želji po čim boljši razumljivosti sintetiziranega govora smo se odločili za prvo možnost, tako da pri izgovarjavi kratic naglašujemo vsak zlog posebej.

3.4 Preverjanje pravilnosti samodejnega določanja izgovarjave besed

Uspešnost samodejnega določanja naglasnega mesta in fonetičnega prepisa besed smo ocenjevali s primerjavo samodejno določenih izgovarjav besed z izgovarjavami iz slovarja 16.000 najbolj pogostih izgovarjav. Z uporabo pravil za določanje naglasnega mesta in pravil za pretvorbo grafemskega niza v fonetični prepis smo samodejno določili izgovarjave besed iz slovarja. Samodejno določen fonetični prepis slovarskih besed smo primerjali z referenčnim fonetičnim prepisom iz slovarja, ki smo ga določili ročno.

Te izgovarjave smo primerjali z ročno določenimi izgovarjavami. V 71 odstotkih primerov je bila izgovarjava besede pravilno določena. Preostalih 29 odstotkov primerov smo razvrstili glede na vrsto napake na naslednji način:

vrsta napake	odstotek nepravilno naglašanih besed [%]
pravilno določen naglašeni glas, vendar napake v fonetičnem prepisu za preostali del besede	2.1
pravilno mesto naglasa, a napačen naglašeni glas (zamenjana široki ozki pri <o> in pri <e>)	12.2

nepravilno mesto naglasa in s tem tudi napake v fonetičnem prepisu za preostali del besede	15.1
skupaj	29.4

Tabela 14: Preizkus pravilnosti samodejnega napovedovanja naglasnega mesta in delovanja pravila za grafemsko-fonetični prepis besedila.

Navedena verjetnost napake pri določanju fonetičnega prepisa besede velja za primer, ko slovar izgovarjav ni upoštevan, sicer je verjetnost napake manjša.

4. PROZODIČNE LASTNOSTI GOVORA

Pravilna izbira prozodičnih parametrov je zelo pomembna za tvorjenje naravnega in razumljivega sintetiziranega govora, zato ji moramo pri načrtovanju sintetizatorja govora posvetiti precej pozornosti.

Osnovne tri prozodične parametre: osnovno frekvenco F_0 , jakost in trajanje, je treba nastaviti v skladu s segmentno strukturo, slovarjem, skladnjo in semantiko sporočila. Pri tem moramo vedeti, kako se v danem jeziku izražata predsem dva prozodična pojava: *grupiranje besedila in naglaševanje*.

Na primer, treba je preveriti, ali so meje med skupinami besed v danem jeziku zaznamovane s spremembo v trajanju, z globalnimi intonacijskimi vzorci, z lokalno spremembo v naglasu, ali pa morda s kombinacijo vseh teh dejavnikov.

Področje prozodije je pomembno tako za jezikoslovce fonetike, računalniške jezikoslovce kot tudi za inženirje razvijalce samodejnih govornih odzivnih sistemov. Zato je mnogo na to tematiko polno izraznih nedoslednosti in dvoumnosti. Deloma je tako tudi zato, ker standardno tehnično izražje za prozodijo še ni usklajeno. Pri Wernerju [Werner 94, str. 38] je zbranih nekaj izrazov, ki se pogosto navajajo v literaturi in ki imajo več pomenov, zato so pogosto povod za nejasnosti.

Do izrazijske zmede pride deloma tudi zato, ker lahko prozodične lastnosti govora opazujemo na štirih ravneh uresničitve, kar ustreza štirim stopnjam v govorni komunikaciji [Werner 94].

Prva raven je jezikovna raven. Ko govorec želi izgovoriti misel, s pomočjo prozodičnih in drugih govornih lastnosti izraža svoj namen. Z jezikovnega stališča je namen prozodije, da omogoča izražanje različnih jezikovnih pojavov, kot so poudarjanje določenih delov besedila in zaznamovanje meja med deli besedila. V zvezi s tem govorimo o *tonu, intonaciji in poudarku* ali *naglasu*. Te pojme bomo natančno predstavili v nadaljevanju poglavja.

Druga raven je raven izgovorjave ali artikulacije. Na tej ravni se prozodične lastnosti govora fizikalno odražajo kot zaporedje sprememb položajev in oblik govoril v govornem traktu. Če jih želimo izmeriti, potrebujemo zahtevno opremo, npr. magnetograf, rentgen ali ultrazvočno merilno napravo.

Tretja raven je raven akustične uresničitve; poglobitvi akustični parametri govornega signala, ki se nanašajo na prozodijo, so *osnovna frekvenca* F_0 , *jakost* in *trajanje*. Naglašeni zlogi imajo npr. pogosto višjo osnovno frekvenco, večjo amplitudo in daljše trajanje kot nenaglašeni zlogi. To ne velja na splošno, kajti pogosto je naglašeni zlog zaznamovan le z dvema od omenjenih treh parametrov. Prav tako se lahko pripeti, da je naglašeni zlog izgovorjen z nižjo osnovno frekvenco kot nenaglašeni zlog.

Četrta raven je raven zaznavanja. Poslušalec prozodijo zaznava kot *premore*, *dolžino*, *melodijo* ali potek višine osnovnega tona oz. preprosto tonski potek in *glasnost*.

Odziv poslušalca na prozodične pojave lahko preverjamo s psihoakustičnimi testi, kjer lahko v akustičnih parametrih izmerimo razliko, ki je potrebna, da pri poslušalcu izzove minimalno spremembo v zaznavi dveh različnih govorčevih namenov. Spremembo višine osnovnega tona na tonski krivulji lahko poslušalec zazna le, če posluša trajanje vsaj 100 ms dolgo in je sprememba višine v tem intervalu vsaj 14 % do 17.5 % osnovne frekvence [Newlekowsky 73]. Najmanjša slušno zaznavna sprememba trajanja se giblje v razponu med 10 in 15 ms [Bakran 84].

V skladu z jezikoslovno tradicijo se pojem prozodija nanaša na govorne značilnosti, ki se ne nanašajo na en sam fonetični segment, glas, temveč na večje enote, ki vključujejo več fonetičnih segmentov, kot so besede, fraze, stavki ali celo večji odseki govorjenega besedila [Werner 94]. Zato o prozodičnih lastnostih govora pogosto govorimo kot o *nadsegmentnih prozodičnih lastnostih*. Tekoči govor strukturirajo v prozodične fraze. Zaznavamo jih kot: naglas, spremembo v intonaciji, ritem in glasnost.

Meritve so pokazale, da na nižjem nivoju na prozodijo vpliva tudi segmentna struktura govora, takrat govorimo o *mikroprozodiji* ali *segmentnih prozodičnih lastnostih* [Werner 94]. Na trajanje in osnovno frekvenco govornega segmenta F_0 vplivajo okoliški glasovi. Tako je npr. v angleškem govorjenem jeziku začetna F_0 samoglasnika, ki sledi zaporniku, nižja od začetne F_0 , ki sledi priporniku. Prav tako je trajanje samoglasnikov v nezvenečem okolju navadno krajše od trajanja samoglasnikov v zvenečem okolju.

Mikroprozodične pojave je v tekočem govoru težko opazovati, ker je pri njihovem merjenju zelo težko izločiti vpliv močno dominantnih višjenivojskih nadsegmentnih pojavov, zato jih navadno opazujemo na ločeno izgovorjenih besedah.

Pri podrobnem opisu prozodičnih pojavov se bomo omejili predvsem na tiste, ki jih lahko določimo in ovrednotimo s postopki obdelave govornega signala, in njihove uresničitve na *jezikovni ravni*: naglaševanje, intonacijo, ritem in hitrost govora.

4.1 Naglaševanje

V tekočem govoru je posamezen zlog lahko izgovorjen glasneje ali tonsko izraziteje kot sosednji zlogi. Ta pojav imenujemo *naglas* ali *poudarek* [Werner 94]. Glede na obseg domene, ki jo opazujemo, razlikujemo tri vrste naglasa: besedni naglas, frazni naglas ali frazni poudarek in stavčni naglas ali stavčni poudarek. Čeprav Werner uporablja besedo poudarek za vse tri kategorije, se bomo v nada-

ljevanju tega dela zgledovali po terminologiji, kot jo predlaga slovnica slovenskega jezika. Tam je zabeleženo, da je naglas značilnost besede, poudarek pa značilnost stavka [Toporišič 91, str. 57].

Čeprav domena, v kateri obravnavamo naglas, lahko zajema besede, fraze, povedi ali stavke, je nosilec naglasa vedno zlog, ki ga imenujemo naglašeni zlog oz. poudarjeni zlog.

Glede na jezik, ki ga obravnavamo, igrata pri akustični uresničitvi naglasa vlogo vsaj dva izmed naslednjih treh akustičnih parametrov prozodije: osnovna frekvenca, jakost in trajanje [Werner 94]. V tonskih jezikih, značilen primer je mandarinska kitajščina, se posamezne besede razlikujejo med seboj le po obliki ovojnice F_0 . V teh primerih melodične opozicije prevzamejo vlogo fonemskih vrednosti, saj se pomen besede ustvari s pomočjo intonacije.

Mnogo jezikov v večzložnih besedah, frazah ali stavkih razlikuje primarni naglas ali primarni poudarek, sekundarni naglas ali sekundarni poudarek in oslabljeni ali reducirani naglas.

V besedah, frazah ali stavkih, ki vsebujejo več naglašanih zlogov, je najbolj izrazit primarni poudarek, sledi mu sekundarni poudarek. Navadno se frazni in stavčni poudarek ujemata z besednim naglasom. Preostale besede oz. zlogi lahko nosijo oslabljeni ali reducirani naglas. Kljub redukciji naglasa vse take besede še vedno nosijo svoj lastni *besedni* naglasni vzorec.

4.2 Intonacija

Drugi pomembnejši prozodični pojav zaznamo kot *melodijo* govora, strokovno jo imenujemo *intonacija* [Werner 94].

Govorci večine indoevropskih jezikov zaznamujejo naklon, kot je npr. povedni naklon nasproti vprašalnemu naklonu, s spremembo melodije pri izgovarjavi povedi. Tako povedi <Danes je lepo vreme.> in <Danes je lepo vreme?> ločimo po razliki v intonacijskih vzorcih: prva intonacija ima padajoč intonacijski vzorec, druga intonacija ima rastoč intonacijski vzorec. Nadalje lahko govorec uporabi intonacijo za izražanje paralingvističnih stališč o vsebini povedanega, npr. za izražanje dvoma.

Akustična manifestacija intonacije se odraža predvsem v osnovni frekvenci F_0 . Za namen sinteze govora je ključnega pomena, da poznamo povezavo med spremembami v osnovni frekvenci, ki so pogojene z naglasom, gl. prejšnji razdelek, ter med tistimi, ki jih pogojuje intonacija.

Najbolj preprosto bi bilo, če bi smeli predpostaviti, da sta oba pojava kumulativna. Žal ni tako. Meritve so pokazale, da je frekvenčni potek besede z istim naglasom v dveh različnih intonacijah lahko povsem različen od vsote vpliva obeh dejavnikov [Werner 94]. Tako npr. vprašalna intonacija ni vedno zaznamovana le z naraščajočo frekvenco v zadnjem naglašeni zlogu, temveč tudi s padcem

frekvence pred takim naglašenim zlogom. Do tega pojava pogosto pride pri večzložnih besedah, ki nimajo naglasa na prvem zlogu.

Intonacije ne moremo predvideti na preprost način. Raziskave zadnjih tridesetih let na tem področju so pokazale, da je potek F_0 zelo zapleten in se od govorca do govorca močno razlikuje, tudi pri enem samem govorcju se manifestira na različne načine. Potek F_0 je odvisen od števila zlogov, položaja primarnega in sekundarnega naglasa in povezanosti z ostalimi intonacijskimi spremenljivkami.

Način določanja intonacije, ki se uporablja v večini sistemov za sintezo govora, je še vedno zelo preprost. Sistemi, ki so sposobni karseda verodostojno modelirati intonacijo, so redki, še bolj redki sistemi vključujejo ustrezne modulacije F_0 na besedni ravni. Na tem področju so prav gotovo potrebne še številne raziskave in meritve.

4.3 Ritem in hitrost govora

Tretji prozodični pojav se nanaša na spremembe v hitrosti tvorjenja govora. To se odraža v dveh vrstah zaznavnih vtisov, ki sta odvisni od dolžine časovne domene, kjer prihaja do takih sprememb. V primeru, ko sta celotna poved ali sestavek izgovarjena hitro, zmerno hitro ali počasi, govorimo o spremembi *ritma* ali spremembi *hitrosti govora*. Če pa je časovna sprememba bolj lokalnega značaja, se časovno spremembo trajanja pogosto pripiše naglaševanju. Lokalna upočasnitev govora, ki se odrazi v povečanem trajanju celotne besede, je značilna za naglaševanje. Četudi je beseda izgovorjena brez izrazitega povečanja glasnosti, izraža beseda, ki je izgovarjana počasneje, poseben pomen. In nasprotno, semantično manj pomembne besede izgovorimo hitreje.

Tako lokalne kot tudi globalne spremembe hitrosti tvorjenja govora niso popolnoma linearne. Zato s preprostim enakomernim upočasnjevanjem ali pohitritvijo ne moremo doseči naravnega govora. Spremembe v hitrosti govora se najbolj močno odrazijo na trajanju samoglasnikov in najmanj na trajanju zapornikov.

4.4 Glasnost

V časovnem poteku govornega signala se *jakost* signala in s tem slušno zaznavna glasnost nenehno spreminjata. Ta pojav imenujemo *dinamika* govornega signala. Do nje pride zaradi različnega načina tvorjenja glasov. Tako so zvoneči glasovi izgovarjani bolj glasno kot nezvoneči. Pri stavčnem poteku glasnosti opazimo, da glasnost proti koncu povednih stavkov navadno upada, medtem ko narašča proti koncu vprašalnih stavkov [Eppinger 86, str. 294].

Nihanja v jakosti in posledično glasnosti so v večji meri naravno vsebovana v

posnetih glasovih in jih zato pri sintezi govora, ki temelji na združevanju osnovnih govornih enot, ni treba dodatno upoštevati [Eppinger 86, str. 294].

4.5 Splošni in jezikovno odvisni prozodični pojavi

Nekateri prozodični pojavi delujejo v različnih jezikih na skoraj povsem enak način, ker temeljijo na nevroloških in fizioloških vidikih tvorjenja govora pri človeku. Zanje pravimo, da so univerzalni.

Preostali prozodični pojavi pa se lahko od jezika do jezika izražajo na povsem različne načine. Naslednja podpoglavja navajajo primere nekaterih splošnih in jezikovno odvisnih prozodičnih pojavov.

Deklinacija in zmanjšanje obsega sodita med splošne prozodične pojave. Znotraj ene vdišne skupine kaže F_0 težnjo stalnega upadanja. Ta pojav se imenuje *deklinacija* in je fiziološkega izvora [Keller 94]. Nastopi kot posledica enakomernega zmanjševanja pritiska v pljučih.

Znotraj vdišne skupine s časom upada tudi spremenljivost osnove frekvence. Pri tem zgornja frekvenčna meja upada bolj hitro kot spodnja frekvenčna meja, ki se pogosto sploh ne spreminja. Ta pojav se imenuje *zmanjševanje obsega* in ima prav tako fiziološki izvor. Ob koncu vdišne skupine je namreč pritisk v pljučih manjši in je zato ustvarjanje višjih frekvenc oteženo.

V spontanem govoru se besede združujejo v skupine po prozodičnih načelih. Tako imajo npr. besede, ki so skladenjsko povezane z naglašeno besedo, nekatere prozodične lastnosti podobne (ritem ali tempo artikulacije).

Tako se npr. dolgi premori, podaljšanje trajanja in padec osnovne frekvence ujamajo s koncem fraze. Take skupine različni avtorji poimenujejo na več načinov: tonska skupina, naglasna skupina, prozodična fraza ali prozodična beseda.

Čeprav se mnogo prozodičnih pojavov v različnih jezikih odrazi na podoben način, obstajajo tudi taki, ki so značilni za posamezne jezike.

Sem se uvršča predvsem način oblikovanja naglasa, postavljanje naglasnega mesta in postavljanje premorov, izraznostna vloga intonacijskih vzorcev in medsebojna povezava med parametri F_0 , jakostjo in trajanjem.

5. PROZODIČNE LASTNOSTI SLOVENSKEGA KNJIŽNEGA JEZIKA

V tem poglavju opisujemo, kako se različni prozodični pojavi odražajo v govorjenem slovenskem knjižnem jeziku. Podajamo pregled rezultatov raziskav, ki so jih opravili slovenski znanstveniki, predvsem jezikoslovci, in katerih izsledke smo delno upoštevali pri nastavljanju prozodičnih parametrov za sintetizirani govor.

Poleg osnovnih (segmentnih) lastnosti, ki izhajajo iz položaja govoril v ustni votlini, imajo glasovi še t. i. nadsegmentne (prozodične) lastnosti, tj. trajanje, jakost in osnovno frekvenco [Toporišič 91, str. 52].

Po dolžini trajanja so samoglasniki lahko dolgi ali kratki, po jakosti krepki ali šibki, po tonu pa visoki (cirkumflektirani) ali nizki (akutirani). Jakostno in tonsko izrazitost samoglasnikov imenujemo *naglas*, izrazitost po trajanju pa *količnost* ali kvantiteta.

Poznavanje zakonitosti v trajanju in višini osnovnega tona za slovenske glasove je skopo. S tega področja je bilo opravljenih malo fonetičnih raziskav. Najbolj sistematična med njimi je študija Srebotove [Srebot 88], ki obravnava trajanje in osnovno frekvenco za slovenske samoglasnike v različnih legah.

Praden opišemo naše meritve in dognanja, podajamo pregled do sedaj opravljenega dela na področju merjenja in razlaganja prozodičnih parametrov v slovenskem knjižnem jeziku. Pregled smo razdelili na tri ločene dele, ki obravnavajo različno velike enote besedila: fonetiko glasov – trajanje in osnovna frekvenca glasov, fonetiko besed – naglaševanje besed in stavčno fonetiko – členitev besedila s premori, poudarjanje, stavčna intonacija, register, hitrost govora in tempo govora.

V tem poglavju bomo govorili o trajanju in osnovni frekvenci posameznih glasov v različnih legah v slovenskem knjižnem jeziku.

5.1 Trajanje glasov v slovenskem knjižnem jeziku

Trajanje posameznih glasov se med seboj znatno razlikuje, še zlasti v tekočem govoru [Omerza 70, str. 105]. Glasove, ki jih lahko več ali manj časa držimo ali vlečemo, imenujejo trajniki [Toporišič 91, str. 52]. To so vsi samoglasniki, zvočniki in priporniki. Vsi ostali glasovi, katerih ne moremo vleči, so netrajniki. Ti so zaporniki in zlitniki.

Trajanje glasu zaznamo kot njegovo dolžino v določeni besedi ali frazi. Tako ima vsak glas v besedi svoje določeno trajanje, katerega ne smemo preveč spreminjati, ker bi s tem porušili ravnovesje v govoru in ga tako napravili manj ra-

zumljivega [Omerza 70, str. 105]. Dolžinska razmerja med glasovi morajo ostati vedno pravšnja, ne glede na to, ali govorimo zmerno, hitro ali počasi.

Poleg ostalih dejavnikov, ki vplivajo na trajanje posameznega glasu, trajanje glasu predpisuje tudi njegova lastna fonetična kakovost, ki jo določata predvsem način in pa mesto artikulacije glasu [Srebot 88, str. XX; Bakran 84, str. 47]. Tako trajata dva različna glasova, ki ju postavimo v enako fonetično okolje, kjer tvorita minimalni par, različno dolgo. Ta razlika v trajanju je pogojena izključno z izgovornimi značilnostmi za posamezen glas. Pravimo, da ima vsak glas svoje značilno *inherentno trajanje*, ki ga po času trajanja razlikuje od drugega glasu v enakem fonetičnem okolju. Kot v drugih jezikih, sta tudi v slovenščini inherentno trajanje in osnovna frekvenca samoglasnikov odvisna od lege jezika pri izgovarjavi posameznih samoglasnikov.

Pri merjenju trajanja glasov se navadno kot inherentno trajanje upošteva povprečna vrednost trajanja za določen glas v določenem merilnem poskusu [Lehiste 72]. To trajanje odraža, poleg lastnega značaja glasu tudi vpliv okoliških glasov, tempa artikulacije, naglašenosti zloga na trajanje tega glasu. Menimo, da se izmerjeno povprečno trajanje glasu približa inherentnemu trajanju glasu v primeru, če imamo na voljo dovolj meritev za posamezen glas. V tem primeru inherentno trajanje ostane odvisno le še od hitrosti govora.

5.2 Trajanje samoglasnikov

Slovenski jezik ne pozna izrazito dolgih samoglasnikov, če jih primerjamo z angleškimi ali s češkimi, ker je trajanje samoglasnikov v slovenščini vezano tudi na osnovno frekvenco [Omerza 70, str. 106]. Trajanje samoglasnikov in njihova osnovna frekvenca sta namreč komplementarna: čim višja je lega jezika, tem krajši je izgovorjeni samoglasnik in višja je njegova osnovna frekvenca; čim nižja je lega jezika, tem daljši je samoglasnik in nižja je njegova osnovna frekvenca [Srebot 88, str. 242].

Splošne ugotovitve o sposobnosti slušnega zaznavanja človeka so pokazale, da poslušalec lahko jasno razpozna ločeno izgovarjan samoglasnik, če ta traja vsaj 30 do 40 ms dolgo [Lehiste 70]. Pri hitrem govoru je trajanje samoglasnikov lahko tudi krajše, le da v takem primeru samoglasnik lahko pravilno prepoznamo le, če je izgovorjen v sklopu z drugimi glasovi.

S sluhom samim ni težko razlikovati več dolžinskih stopenj, vendar za praktične namene zadostujejo dve ali tri stopnje. Ko govorimo o dveh stopnjah, je ena *kratka* in druga *dolga*. Če nam to ne zadostuje, uvedemo še tretjo, srednjo stopnjo.

Po navedbah v slovenski slovnici [Toporišič 91, str. 52] večina slovenskih narečij še razlikuje med dolgimi in kratkimi naglašeni samoglasniki. To pa se ne ujema z rezultati študije Srebotove [Srebot 88], kjer je avtorica z meritvami tra-

janja glasov pri izučeni govornici ugotovila, da razlikovanje med dolgimi naglašeni in kratkimi naglašeni glasovi ni fonološko relevantno, saj se trajanja dolgih in kratkih glasov v parih besed niso razlikovala tako, kot je bilo pričakovano. Še več, ponekod so praviloma kratki glasovi trajali dlje kot njihovi dolgi pari. Do podobnega zaključka so prišli tudi v nedavni študiji [Petek 96].

Če pa kljub temu še vedno želimo razlikovati med dolgimi in kratkimi naglašeni samoglasniki, upoštevamo naslednje napotke. Dolgi so v slovenskem jeziku lahko le naglašeni samoglasniki, nenaglašeni izglasni so lahko le položajno dolgi [Toporišič 91, str. 52]. Posebnosti pri tem so naslednje. Ozka [e:] in [o] sta vedno le dolga in naglašena, razen v redkih primerih, ko se ozki glas pojavi v breznaglasnici, npr. zveze s pomožnim glagolom biti: <bomo naredili> [bomO narEdi:li]. [i] [a] [u] in široka [E] in [O] so lahko dolgi ali kratki in kot kratki so lahko naglašeni ali nenaglašeni. Če so kratki in naglašeni, se navadno nahajajo v zaprtem zlogu. Polglasnik [@] je vedno kratek, ne glede na to, ali je naglašen ali nenaglašen. Naglašena glasovna zveza [@] + [r] je vedno dolga. Nenaglašeni samoglasniki so fonološko vedno kratki.

Časovno trajanje samoglasnikov se razlikuje glede na zgradbo zloga, ali se nahajajo v naglašenem ali nenaglašenem zlogu, ali je zlog odprt ali zaprt, ter glede na splošno lego v stavku, torej ali je zlog končni ali ne. Na trajanje samoglasnika lahko dodatno vpliva tudi fonetični kontekst, predvsem soseščina soglasnikov, s katerimi je samoglasnik obdan [Peterson 60].

Prve instrumentalne meritve trajanja slovenskih samoglasnikov je opravil F. Bezlaj [Bezlaj 39]. Določil je njihovo trajanje pri navadni hitrosti govora. Ugotovil je, da vsi samoglasniki niso enako dolgi. Trajanje dolgih naglašeni samoglasnikov leži v razponu med 103 ms in 172 ms. Nenaglašeni samoglasniki so približno za polovico do dveh tretjin krajši kot naglašeni. Kratka naglašena [i] in [u] sta lahko celo krajša od polglasnika.

Najbolj sistematične meritve trajanja slovenskih samoglasnikov je opravila T. Srebot Rejec in svoje ugotovitve objavila v knjižni obliki [Srebot 88, str. 21–32]. Slovenske samoglasnike je razvrstila po njihovem inherentnem trajanju in dobila naslednji vrstni red za naglašene odprte in zaprte zloge [Srebot 89]. Za knjižno slovenščino sta značilna posebej kratka [i] in [u]. Po trajanju v odprtih zlogih sta [i] in [u] kot najbolj zaprta samoglasnika najkrajša in [a] kot najbolj odprt najdaljši. [e] in [E] sta si zelo blizu in enako [o] in [O]. Dolžinska razlika med členoma enega para je sicer zelo majhna, je pa bolj zaprti člen daljši kot bolj odprti in ne obratno, kot bi bilo pričakovati. Tudi sprednji samoglasniki so daljši v primerjavi z ustreznimi zadnjimi. V zaprtih zlogih so vsi samoglasniki na splošno krajši. Še vedno pa sta [i] in [u] najkrajša in sta skupina zase, sledijo jima sredinski in najdaljši je zopet [a] kot najbolj odprt.

V enakem glasovnem okolju je trajanje samoglasnikov odvisno od višine lege jezika. Zato sta samoglasnika [i] in [u] najkrajša, samoglasnik [a] pa je najdaljši, ker ima najnižjo lego jezika.

Študija Srebotove [Srebot 88, str. 246–247] strne ugotovitve o trajanju slovenskih samoglasnikov v naslednjih točkah. Redko kateri govorec sistematično razlikuje med pričakovanimi dolgimi in kratkimi samoglasniki na oksitonih. Dolžina samoglasnikov v knjižni slovenščini z ljubljansko osnovo ni fonološko relevantna. Najmočnejši vpliv na trajanje samoglasnika ima naglas, nato mu sledi inherentno trajanje; ostali vplivi so manjši (tempo govora, struktura zloga, dolžina besede, položaj besede v stavku in drugo). Kratkih cirkumfleksov v oksitonih skorajda ni. V knjižni slovenščini ni kratkega akuta na polglasniku. Polglasnik je inherentno kratek, če pa je naglašen, je daljši (še zlasti v odprtih zlogih). Če imamo več prednaglasnih zlogov, trajanje prednaglasnih samoglasnikov narašča z bližino naglašeneemu zlogu. Čim daljša je beseda, tem krajši so samoglasniki. Paradigmatska primerjava trajanja naglašeneega samoglasnika v enozložni besedi s trajanjem istega, sedaj prednaglasnega samoglasnika v dvo- in štirizložnih besedah pokaže, da znaša trajanje v dvozložnici prednaglasnega samoglasnika približno 66 % trajanja naglašeneega samoglasnika in v štirizložnici 53 % trajanja naglašeneega samoglasnika.

V Laboratoriju za umetno zaznavanje na Fakulteti za elektrotehniko in računalništvo je bilo že predhodno opravljenih nekaj meritev trajanja in osnovne frekvence slovenskih samoglasnikov v tekočem govoru [Mihelič 93a,b]. Posneli so tri nepoklicne govorko, ki so prebrale predloženo besedilo. Razlikovali so med naglašeni in nenaglašeni samoglasniki, niso pa posebej obravnavali dolgih oziroma kratkih samoglasnikov. Rezultati meritev se precej ujemajo s tistimi, dobljenimi pri Srebotovi [Srebot 88], čeprav je šlo v prvem primeru za tekoč govor. Statistični Student *t*-test je potrdil razlike v časih trajanja med naglašeni in nenaglašeni samoglasniki, pri tem da vrsta naglasa ni bila fonetično ovrednotena. Opazili so izrazito bimodalno porazdelitev pri časih trajanja za samoglasnike [a], [E] in [O], kar opravičuje statistično delitev teh samoglasnikov na dolge in na kratke. To seveda še ne pomeni, da je taka delitev tudi fonološko relevantna. Izmerjene vrednosti so bile precej raztresene, kar tolmačimo kot posledico tekočega govora in nepripravljenosti govork na branje izbranega besedila.

Zadnje objavljeno delo, ki govori o trajanju slovenskih samoglasnikov [Petek 96], dodatno potrjuje ugotovitev, podano pri Srebotovi [Srebot 88], da slovenski govorniki ne delajo razlike med kratkimi in dolgimi naglašeni samoglasniki. Kot morebitno izjemo navajajo samoglasnik [a].

5.3 Trajanje soglasnikov

Na trajanje soglasnikov vpliva mnogo dejavnikov [Srebot 88, str. 19–32]: vrsta zloga, v katerem se soglasnik nahaja (naglašen ali nenaglašen); število soglasnikov v zlogu (kopičenje soglasnikov v skupini vpliva na njihovo krajše trajanje)

in s tem tudi glasovno okolje, v katerem se soglasnik nahaja (sklop VCV (samoglasnik-soglasnik-samoglasnik) ali sklop CC (soglasnik-soglasnik)); položaj zloga v besedi in položaj besed v stavku (trajanje zadnjega zloga v besedi pred preomorom je navadno daljše od trajanja preostalih zlogov).

Trajanje slovenskih soglasnikov je prvi instrumentalno meril dr. F. Bezljaj [Bezljaj 39]. Meritve je opravil na več govorcih in meril povprečne vrednosti trajanja soglasnikov v naglašeni in v nenaglašeni zlogi. Njegovi rezultati so bili presenetljivi, saj se pričakovanje, da bodo zvoneči soglasniki daljši od nezvonečih, ni potrdilo.

Podobne meritve so bile opravljene za hrvaški jezik [Bakran 84]. Bakran je proučeval trajanje soglasnikov v dveh sklopih, in sicer v primerih, ko se soglasnik nahaja v VCV sklopu, ter v primerih, ko se soglasnik nahaja v sklopih CC.

5.4 Osnovna frekvenca samoglasnikov

Osnovne frekvence slovenskih samoglasnikov se, prav tako kot trajanja, razlikujejo glede na to, ali se samoglasnik nahaja v naglašeni ali nenaglašeni zlogi, odvisne pa so tudi od splošne lege v stavku, torej ali je zlog odprt ali zaprt in ali je končni zlog ali ne. Značilno je, da so osnovne frekvence samoglasnikov v odprtih zlogih znatno nižje od tistih v zaprtih, saj so odprti zlogi daljši od zaprtih [Srebot 88].

Osnovna frekvenca samoglasnika v naglašeni zlogi je odvisna tudi od vrste naglasa. Tonsko območje obeh vrst tonemskih naglasov pri posameznem govorcu teži za tem, da je za oba naglasa enako [Srebot 88, str. 246].

Podobno kot glasu pripada inherentno trajanje, ki je glasu lastno in ga določa način artikulacije tega glasu, ima glas tudi *inherentno osnovno frekvenco*. Inherentna osnovna frekvenca se od govorca do govorca močno razlikuje. Inherentne osnovne frekvence za posamezne samoglasnike si sledijo podobno kot inherentna trajanja: zaprti samoglasniki, sredinski, odprti, le da sta tu odprta člena sredinskega para v skupini s samoglasnikom [a] in ne s svojim zaprtim parom. V odprtih zlogih so tri skupine: [i], [u], zaprta sredinska samoglasnika [e], [o] in odprti samoglasniki [O], [E] in [a]. Frekvenčni razpon v zaprtih zlogih bi bil znatno manjši kot v odprtih, če ne bi izstopal [u] s svojo visoko frekvenco.

Na že omenjenih meritvah, opravljenih na Fakulteti za elektrotehniko in računalništvo [Mihelič 93a,b] so bile poleg meritev trajanja izvedene tudi meritve osnovne frekvence slovenskih samoglasnikov. Statistični test Student *t*-test je pokazal, da razlike med povprečnimi osnovnimi frekvencami med naglašeni in nenaglašeni samoglasniki niso merodajne. Domnevamo, da je razlog v tem, ker se ni razlikovalo med različnimi vrstami tonemskih naglasov. Možno je, da so govorke govorile z akutnim naglasom, kjer tonski vrh še ni dosežen na naglašeni samoglasniku, ampak šele na ponaglasni zlogi.

Pri zaznavanju tonskega poteka govora osnovna frekvenca zvonečih soglasnikov ni tako pomembna, ker soglasniki niso pglavlitni nosilci slušno zaznavnih informacij.

Pri zvočnikih v končni legi je osnovna frekvenca nekoliko višja, ker takrat zvočnik ni omejen z naslednjim soglasnikom, in lahko izzveni.

5.5 Besedni naglas v slovenskem knjižnem jeziku

V slovenskem knjižnem jeziku besede naglašujemo na dva načina, jakostno in tonemsko [Toporišič 91, str. 55], [SAZU 90, str. 75]. *Jakostni naglas* pomeni, da bo označeni glas izgovarjan z močnejšim glasom, navadno je tak glas tudi tonsko višji. *Tonemski naglas* določa, da so samoglasniki enkrat izgovarjani više, drugič pa nižje.

Tonemsko naglašeni samoglasniki so izgovarjani v sicer enakem glasovnem sestavu enkrat višje, cirkumfektirani, drugič nižje, akutirani; po stari terminologiji so prvi padajoči, drugi rastoči. Oba tipa naglasa pri tonemskem naglaševanju imenujemo *tonema*. Tako poznamo tonem *akut* in tonem *cirkumfleks*.

Obe vrsti naglaševanja sta enakovredni, vendar večinoma pokrajinsko vezani, tako je tonemsko naglaševanje značilno npr. za dolensko, gorenjsko, koroško in ljubljansko govorno področje [SAZU 90, str. 75]. Tonemskega naglasa se naučimo v mladosti in je narečno pogojen. Slovar slovenskega knjižnega jezika podaja pri geslih tudi vrsto tonemskega naglasa [SSKJ 94]. Pri nekaterih slovarskih geslih dopušča več možnih naglasov. Jezikoslovci so pri določanju tonemskega naglasa opazili dve pravili. Prvič, v oksitonih je cirkumfleksni naglas mnogo bolj pogost kot akutni [Srebot 88, str. 236] in drugič, da je pri novejših izposojenkah cirkumfleksni naglas bolj pogost [Rigler 68, str. 198].

Vloga jakostnega naglasa je lahko pomenskorazločevalna. Loči med posameznimi besednimi vrstami, npr. med predlogom <čez> in časovnim prislovom <čez> [tSe:s] proti [tSEs]. Kadar je v besedi več zlogov, je lahko pomenskorazločevalno tudi samo mesto naglasa, npr. velelnik in 3. oseba ednine v povednem naklonu glagola <oblači> [Obla:tSi] proti [OblatSi:]. Pri tonemskem govoru je vloga akuta ali cirkumfleksa v glavnem oblikorazločevalna in le v redkih primerih pomenskorazločevalna [Toporišič 91, str. 57].

Glede na položaj zloga proti naglašenemu zlogu razlikujemo: naglašeni zlog (zlog, ki ima jakostno ali tonemsko izrazitost), prednaglasni zlog (zlog ali zlogi pred naglasom) in ponaglasni zlog (zlog ali zlogi po naglašenem zlogu).

Samoglasniki so glasovi največje odprtostne stopnje [SAZU 90, str. 73]. Glede na dvig jezične ploskve razlikujemo visoke, sredinske in nizke samoglasnike; glede na to, kateremu delu neba se jezik najbolj približa, pa sprednje, srednje in zadnje. Naglašeni so lahko samoglasniki in soglasnik [r]. Glede na naglašenost

oziroma nenaglašenost in glede na dolgost ali kratkost ločimo v slovenskem knjižnem jeziku tri samoglasniške sestave [SAZU 90, str. 74].

5.6 Tonemski besedni naglas

Začetki raziskav slovenskega tonemskega naglaševanje so izčrpno opisani pri Srebotovi [Srebot 88, str. 1–9], s poudarkom na Bezlajevih in Toporišičevih študijah [Bezljaj 39] in [Toporišič 68]. Najbolj obsežno raziskavo je izvedla T. Srebot Rejec, ki je svoje izsledke podala v knjigi [Srebot 88]. Ker smo si pri določanju poteka osnovne frekvence pogosto pomagali z njenimi dognanji, jih bomo v nadaljevanju poglavja na kratko predstavili.

Za razliko od mnogih drugih jezikov v slovenščini naglasno mesto in vrh tonske krivulje v besedi ne sovpadata vedno [Srebot 88, str. 15]. Zato T. Srebot Rejec slovenščino uvršča med t. i. *pitch accent languages*. Ker tonemski naglas v slovenščini večinoma ni pomenskorazločevalen, uvršča R. Šuštaršič slovenščino, za razliko od tonskih jezikov, med intonacijske jezike [Šuštaršič 94, str. 43].

Vrh tonske krivulje in naglasno mesto ne sovpadeta, ker tonska krivulja doseže svoj vrh šele na ponaglasnem zlogu, čeprav predhodni zlog nedvomno zaznamo kot naglašena [Srebot 88, str. 229]. Če je v besedi naglašen nekončni zlog, je za oblikovanje besednega naglasa pomemben tudi naslednji, ponaglasni zlog. Fonološka značilnost, ki loči oba tonemska naglasa, je razlika v osnovni frekvenci med koncem naglašena in začetkom ponaglasnega zloga. To razliko imenujemo *skok*. Za akut je značilen skok navzgor, za cirkumfleks skok navzdol.

Oksitoni ali besede, ki so naglašene za zadnjem zlogu, nimajo ponaglasnega zloga, in tako tudi skoka ni. Tonska krivulja oksitona je tako lahko naraščajoča (R rise), naraščajoča padajoča (RF rise fall), padajoča (F fall) ali ravna (L level). V slovenščini je za naglas značilno naraščanje tona, torej je osnovna krivulja tona za besedni naglas naraščanje. Čim bolj poudarjena je beseda, tem bolj izrazito in strmo je to naraščanje znotraj naglašena zloga. Po dosegu tonskega vrha sledi upadanje, ki pa je odvisno od vrste besede (bariton ali oksiton); ter od vrste naglasa (akut ali cirkumfleks). Pod vplivom stavčne intonacije pri oksitoni v rezultatih raziskav tega padanja največkrat ni bilo opaziti [Srebot 88]. Pri baritonih pa je to padanje skoraj vedno zaznavno.

Akut in cirkumfleks se v baritonu razlikujeta po tem, kdaj se pojavi tonski vrh: pri cirkumfleksu tonski vrh sovpada z naglašenim zlogom, pri akutu se tonski vrh pojavi šele na ponaglasnem zlogu [Srebot 88, str. 229]. Cirkumfleks baritonski naglas ima hitrejši in strmejši vzpon, medtem ko ima akut bariton počasnejši in položnejši vzpon. Na naglašenem samoglasniku pri cirkumfleksu je ton opazno višji kot pri akutu, ker se pri akutu pojavi tonski vrh šele na ponaglasnem zlogu.

Naraščanje ali padanje se lahko zvezno brez prekinitve nadaljuje na ponagla-

snem zlogu le v primeru, če leži med obema zlogoma zveneči soglasnik. Nezveneči soglasnik, posebno je to izrazito pri zaporniku, na kratko prekine tonsko krivuljo, zatem pa se pojavi na naslednjem glasu na višji (pri akutu) oziroma na nižji (pri cirkumfleksu) točki. To spremembo osnovne frekvence imenujemo *skok*. V primeru, da je naglašeni samoglasnik po svojem trajanju dovolj dolg, ni več potrebe po strmem naraščanju do tonskega vrha. To je značilno predvsem za akutni naglas, kjer naraščanje tudi sicer ni tako strmo, saj je tonski vrh dosežen šele na ponaglasnem zlogu. Zato lahko na začetku ton celo malo pade, preden začne naraščati. To pa privede do t. i. konkavne oblike tonske krivulje [Srebot 88, str. 243]. Značilnost take krivulje je ponavadi negativni *tonski premik* (angl. pitch movement) znotraj naglašenege zloga, ki podaja razliko med osnovno frekvenco na začetku in na koncu zloga.

Negativni tonski premik pomeni, da je na koncu zloga ton nižji kot na začetku zloga. Ponavadi sledi v takem primeru precej velik skok navzgor na ponaglasnem zlogu. Izkazalo se je, da je padanje, ki sledi na ponaglasnem zlogu, manj strmo (ali pa gre celo za rahlo naraščanje), tako da se ne izniči predhodni precej veliki skok navzgor. Take oblike tonske krivulje so značilne za bariton akut [Srebot 88]. Zakasneni začetek naraščanja frekvence s predhodnim padanjem ali brez pri bariton akutu pri večini govorcev pomaga obdržati tonski vrh znotraj njihovega frekvenčnega razpona.

V primeru nezvenečega soglasnika, še zlasti pri zaporniku, se pojavi skok navzdol tudi pri bariton cirkumfleksu. Tonski vrh je dosežen na naglašnem zlogu, največkrat proti koncu, bolj redko se ta točka premakne proti sredini naglašenege zloga [Srebot 88].

V nadaljevanju opisujemo zanimive rezultate preverjanja tonemskega naglasa s slušno zaznavnim tolmačenjem izgovarjane besede, ki so podrobno opisani pri Srebotovi [Srebot 88]. Trije izbrani poklicni govorci so izgovarjali besede z naglasom, ki se je v večini primerov ujemal s pričakovanim naglasom, kot ga določa [SSKJ 94]. Naglase sta slušno zaznavno tolmačila dva fonetično izobrazena jezikoslovca. Če sta se akustična klasifikacija in percepcijsko tolmačenje ujemala, so take realizacije označili kot značilne ali tipične.

Zanimivi so primeri, ko je prišlo do razhajanja mnenj obeh ocenjevalcev. Vzrok temu so pripisali ne dovolj izrazito izgovarjanim besedam ali pa nekaterim lastnostnim, ki so skupne obema naglasoma. Največkrat so to premajhni skoki, predvsem navzdol pri cirkumfleksih, ali pa krivulja akuta in tonska višina cirkumfleksa. V takem primeru sta tolmača interpretirala vrsto naglasa po tem, čemur sta pripisala večjo pomembnost – višini tona ali obliki tonske krivulje, ali pa po tem, kakšen naglas je bil pričakovani. Pričakovani naglas ima lahko v dvomljivih primerih pomembno vlogo. Če pa je vzorec osnovne frekvence značilen, ni pričakovani naglas prav nič pomemben.

Besede s padajočo ali pa ravno tonsko krivuljo so imele največ neenotnih inter-

pretacij. Vseskozi padajoča ali ravna krivulja v nevtralni intonacijski legi pomeni prešibko naglašen ali prekratek samoglasnik.

Namen raziskave je bil dvojni: proučiti besedni naglas v nevtralni intonacijski legi in določiti značilne vzorce osnovne frekvenca za oba tonemska naglasa [Srebot 88, str. 55–132] ter proučiti vpliv stavčne intonacije in stavčnega poudarka na zaznavanje tonemskih naglasov [Srebot 88, str. 133–209].

5.6.1 Besedni naglas v nevtralni intonacijski legi

Tonemski naglas je najlažje proučevati na besedah, ki se nahajajo v nevtralni intonacijski legi. Opazovane besede so levo in desno obdali z drugimi besedami, tako da je bil vpliv stavčne intonacije na besedo v sredini intonacijskega segmenta majhen [Srebot 88, str. 20].

Pri pregledovanju krivulj osnovne frekvenca so ugotovili, da se v primeru baritonov poteki osnovne frekvenca pri akutu in cirkumfleksu ponavljajo. Tako so sestavili zbirko *značilnih stiliziranih krivulj* osnovne frekvenca za oba naglasa in jih poimenovali *vzorci osnovne frekvenca* [Srebot 88, str. 243].

Ni mogoče trditi, da ti poteki tonskih krivulj zadostujejo za vse besede. To so posplošene najbolj pogoste krivulje, lahko pa se razlikujejo od govorca do govorca ter v odvisnosti od besed (oziroma njihovega poudarka). Npr., četrta skupina cirkumfleksnih tonskih krivulj ne ustreza več značilnemu cirkumfleksnemu naglasu. Izkazalo se je, da je tako izgovarjana beseda razpoznavna kot cirkumfleks samo v primeru, če je prisoten dovolj velik skok navzdol, sicer je razpoznavna kot beseda z akutnim naglasom.

Sledi nekaj splošnih ugotovitev, ki jih lahko povzamemo iz pregleda posameznih tonemskih naglasov. Pri baritonih je naglas lažje razbrati, ker je razporejen na dva zloga, medtem ko se pri oksitonih vse dogaja na enem zlogu, ki nima skoka in je navadno še pod vplivom stavčne intonacije [Srebot 88, str. 244].

Pri govorcih, ki izgovarjajo besede s tipičnimi tonskimi krivuljami in dovolj izrazitimi skoki, so slušno zaznavne interpretacije skoraj vedno enotne. Najbolj tonematično zvenijo govorci, usmerjeni v akut, ker je akutni vzorec osnovne frekvenca lažje identificirati kot cirkumfektiranega. Skok navzgor je namreč lažje slušno zaznati kot enako velik skok navzdol [Srebot 88, str. 236]. Poleg tega je za take govorce značilno, da se besedni naglas močno upira vplivom stavčne intonacije. Če v besedi nadomestimo cirkumfleks z akutom, je to opazno, medtem ko cirkumfektirana beseda namesto akutirane ne moti preveč [Rigler 68, str. 196].

Vse omenjene značilnosti se od govorca do govorca razlikujejo. Nekateri so usmerjeni v izgovarjavo akutnega naglasa, ta prevladuje tudi pri besedah s pričakovanim cirkumfleksnim naglasom, nekateri pa ravno obratno. Za posamezne

govorce je značilno, da določenih naglasov sploh ne izgovarjajo (npr. oksitononov z akutnim naglasom).

Prav tako se pri bariton cirkumfleksih spreminja časovna točka dosega tonskega vrha znotraj naglašene zloga. Nekateri jo dosežejo prej, nekateri pa šele na samem koncu naglašene zloga.

5.6.2 Besedni naglas pod vplivom stavčne intonacije

Pri besedah, ki niso dovolj naglašene, lahko tonska krivulja pada znotraj trajanja celotne besede [Srebot 88, str. 232]. To je posebno izraženo pri padajoči stavčni intonaciji (pred končnimi ločili: pika, podpičje). T. Srebot Rejec je proučevala, kako se tonemski naglas besed podreja vplivom stavčne intonacije [Srebot 88, str. 133–209].

Na besedni naglas vpliva tudi lega besede v povedi. Opazovala je tonemske naglase besed na različnih položajih v povedi oz. v intonacijskem segmentu, pri tem je upoštevala tudi stopnjo jakostnega naglasa na posamezni besedi: na začetku intonacijskega segmenta, sredi intonacijskega segmenta (v nevtralni intonacijski legi), v jedru intonacijskega segmenta, oslajeno naglašene besede, normalno naglašene besede, poudarjeno ali emfatično naglašene besede.

Ugotovitve, zbrane pri Srebotovi [Srebot 88, str. 232 in 245], lahko strnemo v naslednje točke. Osnovna frekvenca pada le v besedah, ki niso naglašene ali pa se nahajajo v padajočem intonacijskem jedru. Pri baritonih lahko vpliv stavčne intonacije nastopi šele na ponaglasnem zlogu in ne že na naglasnem, kar pa je odvisno od govorca in zasidranosti njegovega besednega naglasa. Predvsem pri oksitonih, ki so močno pod vplivom stavčne intonacije in besednega poudarka, ostane pogosto od prvotne tonske krivulje le višina, kontura pa se zabriše, kar je močno odvisno od posameznega govorca. Začetna lega v povedi pospešuje akutni naglas, prav tako pospešuje akut naraščajoča stavčna intonacija, medtem ko padajoča pospešuje cirkumfleksni naglas. Nekateri govorce izgovarjajo pri padajoči stavčni intonaciji skoraj vedno cirkumflekse, pri naraščajoči pa akute, ne glede na pričakovani tonemski naglas izgovorjenih besed, podan v [SSKJ 94]. Pri besedah z emfatičnim poudarkom je besedni naglas še bolj izrazit kot pri besedah v nevtralni intonacijski legi.

Vpliv stavčne intonacije na tonemski naglas je najbolj izrazit na intonacijskem jedru, bodisi naraščajočem ali padajočem, manj izrazit je na začetku intonacijskega segmenta in najmanj vpliven v sredi intonacijskega segmenta [Srebot 88, str. 245]. Vloga besednega naglasa pod emfatičnim poudarkom je okrepljena. Sredi intonacijskega segmenta naglašene besede v veliki meri ohranijo značilnosti svojega tonemskega naglasa.

5.7 Stavčna fonetika

Govorjeno besedilo ni vseskozi izgovarjano neprekinjeno ali enolično: prekinjajo ga premori, melodično je razgibano, določeni deli so izgovarjani glasneje ali hitreje. S tem se ukvarja poseben del fonetike, fonetika besedila ali kar *stavčna fonetika* [Toporišič 91, str. 437]. Pomenske enote v povedi so ločene z daljšimi ali krajšimi prekinitvami, ki jih v pisavi pogosto zaznamujejo ločila [Omerza 70, str. 116]. Tako je govor razdeljen na različne segmente, ki nastopajo kot posledica različnih stavčnofonetičnih pojavov. Segment govora, obdan s premori, nastopi kot posledica členitve govora s premori. Intonacijski segment je govorni segment, ki nosi samostojno intonacijo. Ritmični segment je govorni segment, ki nastopi kot posledica ritmičnosti govora.

Pogosto se naštetih govornih segmenti med seboj pokrijejo. V okviru fonetike besedila govorimo še o: poudarjanju besedila – jakostni izoblikovanosti posameznih govornih segmentov, hitrosti in tempu govora – relativni dolžini trajanja posameznih govornih segmentov in o registru – relativni višinski legi tonskih potev.

5.7.1 Členitev besedila s premori

Členitev besedila s premori je pri govorjenju zelo pomembna. Pravilna členitev sporočilo naredi jasno, pregledno in razumljivo. V nasprotnem primeru je besedilo lahko dvoumno in nerazumljivo.

Premor je stavčnofonetični pojav, ki nastopi, ko se govorna veriga na določenih mestih pretrga. S premori ločeni deli povedi se imenujejo segmenti. Segmenti, ki se oblikujejo med premori, se pogosto ujemajo s preostalimi kategorijami segmentov v okviru stavčne fonetike besedila, zlasti z ritmičnimi in intonacijskimi. Slovnica slovenskega jezika podaja značilna mesta členitve besedila s premori [Toporišič 91, str. 442]. Konci povedi so v pisavi zaznamovani s končnimi ločili (pika, vprašaj, klicaj in tri pike), premori se pojavljajo tudi na meji med stavki iste povedi, zlasti v priredju (najpogosteje v protivnem in v posledičnem, manj dosledno v vezalnem priredju); med ritmičnimi deli povedi (načelo ritmičnosti deli izgovarjano poved na približno enako dolge dele); pri manjši hitrosti govora členitev pogosto nastopi med izhodiščem in jedrom povedi, med vrivkom, pristavkom, dostavkom, izpostavkom, pastavkom in preostalim delom povedi ali stavka, po spremnem stavku premega govora, če je pred dobesednim navedkom, v podredju pred prilastkovim, časovnim in predmetnim odvisnikom, med deli členitve po aktualnosti: delitev stavka na izhodišče ali temo, prehod ali tranzicijo in jedro ali remo [Toporišič 92, str. 17], pred popravki, pri jecljanju, pri čustvenem sporočanju ipd.

Strnjeno se vedno izgovarjajo razvijajoči se stavčni členi glede na jedro: pridev-

nik in samostalnik (prim. <dober človek>), samostalniško jedro in samostalniški prilastek (prim. <znanec iz Ljubljane>), prislov in glagol (prim. <lepo govori-ti>), prislov in pridevnik (<zelo dober>) ter prislov in prislov (prim. <zelo do-bro>).

Premori, ki so vezani na končna ločila, so daljši. Krajši premori so vezani na pristavke in dostavke. Znotraj večstavčne povedi so premori različni glede na vrsto priredja oziroma podredja. Najmanjši so pri vezalnem priredju in največji pri vzročnem priredju in podredju ter pri pojasnjevalnem priredju. Še zlasti majhni, skoraj neobstoječi, so premori med deli nestavčnega priredja z vezniki in, pa, ter: na primer <sestra in brat>. Premor med naslonko in besedo, na katero se naslonka naslanja, se ne pojavi skoraj nikoli.

Naštetih premori, ki nastanejo na značilnih mestih, so različno dolgi, znotraj po-samezne skupine pa imajo značilno trajanje. Na splošno velja pravilo, da so pre-mori tem daljši, čim bolj redno se pojavljajo v besedilih, npr. na koncu povedi.

mesto členitve	trajanje premora [ms]
konec vprašalne povedi	800 do 860
konec pripovedne povedi	700
konec vzklične povedi	580 do 640
konec povedi, ki bi se lahko nadaljevala	580 do 640
konec daljšega dela zložene povedi	350 do 470
meja vrinjenega stavka	420
pred pristavkom, pred veznikom <zato>	270 do 350
med deli priredja (z izjemo veznikov <in>, <ter>)	180 do 240
pred podrednim stavkom	180 do 240
pred zamolkom	180 do 240
pred veznikoma <in>, <ter>	120
pred podrednim veznikom	30 do 70

Tabela 15. Trajanje premorov na različnih mestih v povedi v slovenskem besedilu [Toporišič 91, str. 443].

Toporišič je analiziral trajanje premorov na primeru odlomka iz Cankarjeve črtice, ki ga je bralo sedem različnih govorcev [Toporišič 91, str. 438]. Razlike med premori, ki so jih delali posamezni govorcev na določenih mestih v besedilu, so bile zelo velike, tudi pet- in večkratne. Povprečne vrednosti trajanja premorov na različnih mestih v povedi so zbrane v tabeli 15.

Trajanje premorov v slovenskem besedilu je meril tudi Dobnikar. V članku [Dobnikar 96] navaja, da je analiziral trajanje premorov na raznih mestih členitve na govorni bazi, ki je zajemala različne zvrsti besedila. Premore je združil v štiri skupine in namesto srednje vrednosti za vsako skupino računa mediano. Rezul-tati njegovih meritev se v glavnem ujemajo z vrednostmi, podanimi v tabeli 15.

Poleg besednih naglasov ima lahko vsak stavek tudi enega ali več stavčnih naglasov ali *stavčnih poudarkov*, s katerimi poudarjene besede izgovorimo bolj podkrepljeno kot druge [Omerza 70, str. 117]. V slovnici slovenskega jezika je zapisano, da v povedi lahko razlikujemo štiri stopnje jakostne izrazitosti [Toporišič 91, str. 446]: nenaglašenost, oslABLJENO naglašenost, neoslABLJENO ali normalno naglašenost in poudarjeno ali emfatično naglašenost ali poudarjenost. Stavčni poudarek je subjektiven in nima točno določenega mesta. Odvisen je od tega, kaj v stavku želimo poudariti. Lahko stoji na katerikoli besedi v stavku oz. povedi.

Primerna jakostna členitev je pri branju besedila zelo pomembna, ker je od nje odvisen potek stavčne intonacije. V vsakem segmentu moramo znati odkriti jakostno težišče, ker se v njem oblikuje glava stavčne intonacije. Pri določanju mesta jakostnega težišča veljajo naslednja pravila [Toporišič 91, str. 448]. Če vsebuje segment poudarjeno besedo, leži jakostno težišče na poudarjenem zlogu; jakostno težišče segmentov se nahaja v pomensko najbolj pomembnem delu segmenta. Če noben zlog ni poudarjen, se jakostno težišče nahaja na zadnjem neoslabljenem naglašenem zlogu segmenta. V stavkih z vprašalnim zaimkom ali prislovom je jakostno težišče na naglašenem zlogu zaimka (<zakaj>, <čemu>, <kdaj>, <kdo>, <kako>). Jakostno težišče se lahko nahaja tudi na členku (<ne le iz navade>).

5.7.2 Stavčna intonacija v slovenskem knjižnem jeziku

Stavčna intonacija ali samo intonacija je značilna podoba tonskega poteka oz. poteka višine osnovnega tona v segmentu govora [Toporišič 91, str. 449], ki ga imenujemo intonacijski segment. Poved ima lahko enega ali več intonacijskih segmentov [SAZU 90, str. 57].

Stavčna intonacija ima svoje značilne oblike, ki so pomensko razločevalne. Povedo nam, ali je poved pripovedna ali vprašalna, velelna, želelna ali vzklična: <Danes je lepo vreme.>, <Danes je lepo vreme?>, <Danes je lepo vreme!>.

Potek osnovne frekvence pod vplivom stavčne intonacije za slovenski jezik je bil v zadnjem času pogosto obravnavan. O njem govorijo mnogoštevilne študije [Toporišič 67; Toporišič 78], deloma [Srebot 88; Šuštaršič 94; Vitez 95; Rakar 95; Komar 96] in [Dobnikar 96]. Še vedno pa je bilo izvedenih premalo sistematskih meritev na zadostnem številu akustičnih realizacij slovenskih povedi, da bi lahko sklepali na splošne zakonitosti obnašanja poteka osnovne frekvence pod vplivom stavčne intonacije.

Srebotova je opazovala obnašanje tonemskega naglasa pod vplivom stavčne intonacije [Srebot 88, str. 133–209].

Šuštaršič je prvi opravil primerjalno analizo slovenske in angleške stavčne intonacije [Šuštaršič 94]. Posnel je 55 študentov iz različnih narečnih področij, ki so

izgovarjali dialoge, in slušno zaznavno označeval smer naraščanja oz. padanja višine osnovnega tona. Dobljene intonacijske poteke je primerjal z angleškimi, ki jih je opisal z O'Connorjevim in Arnoldovim modelom [O'Connor 73]. Ugotovil je, da pri večini narečno izbranih govorcev prevladuje netonemski govor. S. Komar je preučevala, kako intonacija vpliva na pomen in sporočilno vrednost izrečenega, še zlasti v govorni interakciji [Komar 96]. V ta namen je opazovala intonacijske poteke na posnetem radijskem intervjuju, tiskovni konferenci in okrogli mizi. Slovensko stavčno intonacijo je opisovala z Brazilovim intonacijskim modelom.

Zanimivi sta študiji Viteza [Vitez 95], ki nakazuje možnost sistematskega opisa podobe poteka višine osnovnega tona za slovenski jezik, ki se bo zgledoval po podobnem opisu za francoski jezik [Auberge 91]; in Dobnikarja [Dobnikar 96], ki predlaga način napovedovanja različnih tipov slovenske stavčne intonacije s Fujisakijevim superpozicijskim modelom [Fujisaki 93].

V nadaljevanju podpoglavja podajamo pogled J. Toporišiča na slovensko stavčno intonacijo, kot ga je opisal v nekaj razpravah [Toporišič 67; Toporišič 78] in v slovnici slovenskega knjižnega jezika [Toporišič 91].

Intonacijski segment sestavljata intonacijsko ogrodje in zložje [Toporišič 91, str. 449]. Podobno opisuje intonacijski segment za angleščino tudi Halliday [Halliday 70]. *Zložje* sestavljajo nenaglašeni zlogi za in pred naglašnim zlogom. Prvo zložje se imenuje zanaglasno, drugo prednaglasno. Zlogi zložja se ne izgovarjajo vsi na isti višini. Naglašeni zlog usmerja zložje v različno stavčno intonacijo, v tonemskih jezikih pa še v različni tonemski naglas: akut ali cirkumfleks. *Intonacijsko ogrodje* sestavljajo naglašeni in poudarjeni zlogi. Če v povedi ni nič poudarjeno, so naglašeni zlogi višinsko urejeni tako, da je prvi najvišji, vsak naslednji pa od njega nižji.

Za bolj pregleden zapis in branje stavčne intonacije, jo zapisujemo v črtovje, podobno notnemu, tako kot pri Toporišiču [Toporišič 91, str. 452]. Navadno uporabljamo štiri vodoravne črte: s tem pridobimo tri pasove govorcevega tonskega obsega. Po potrebi lahko dodamo še četrti pas. Pasovi predstavljajo intonacijski obseg v srednji govorni legi. Tak način zapisovanja intonacije ni absoluten, temveč je relativen, informativen. Izraža relativno tonsko lego med višinami zlogov v intonacijskem segmentu. Naglašene zloge zaznamujemo s črtico <->, nenaglašene s piko <.>.

Intonacijski segment se deli na dva dela, na razločevalno pomembno glavo in nerazločevalno telo ali trup [Toporišič 91, str. 453]. *Intonacijska glava* je tisti del intonacije, ki se začneja z jakostnim težiščem segmenta in traja vse do konca segmenta. *Telo intonacije* je del segmenta, ki se nahaja pred glavo intonacije.

Vsak intonacijski segment ima intonacijsko glavo, telo pa ni nujno treba; nimajo ga npr. segmenti z jakostnim težiščem na prvem naglašnem zlogu.

Poznamo tri oblike glave intonacije: kadenco, antikadenco in polkadenco, glede na skladijski naklon segmenta, ki ga zaključuje: zaključek pripovedne intona-

cije (kadenca), vprašalne intonacije (antikadenca) ali pa nekončnega segmenta (polkadenca). Kadenca in antikadenca sta končni glavi in ju pisno zaznamujemo s končnim ločilom, tj. s piko, z vprašajem ali s klicajem. Polkadenca je nekončna intonacijska glava in jo pisno zaznamujemo z nekončnimi ločili, zlasti z vejico, podpičjem ali dvopičjem; pri vrinjenih stavkih z vprašajem ali s klicajem. Naglašeni zlog, s katerim se začne glava intonacije, imenujemo *težišče intonacije*, sledi mu *rep glave*. Rep glave sestoji ali iz ponaglasnega zložja ali iz zložja in celih taktov.

Tako imenovana členitev po aktualnosti nas uči, da se kadenca, antikadenca in polkadenca začenjajo z naglašnim zlogom besede, ki tvori jedro segmenta sporočila. To jedro sporočila se pri vprašalnih stavkih z zaimensko vprašalnico (npr. <kdo>, <kam>, <zakaj>) nahaja na tej vprašalnici, pri navadnih izjavnih stavkih pa na naglašnem zlogu zadnje besede (gl. mesto jakostnega težišča v poglavju). Kadar smo v dvomih, kje je jedro segmenta, segment izgovorimo z vprašalno intonacijo, in tam, kjer se potek intonacije divergentno preusmeri, je začetek kadence oz. antikadence ali polkadence.

Vzklično intonacijo dobimo z intervalno transformacijo pripovedne ali vprašalne intonacije. To dosežemo s povečanjem intervalnega obsega, tako da težišče intonacije pomaknemo v višje frekvenčno območje.

Polkadenca predstavlja nekončni zaključek intonacijskega segmenta; na primer: <Obljubil je, da bo prišel.>, <Jaz gledam, gledam, poslušam, ni bilo nič.>. Navadno je podobna antikadenci z manjšim intervalnim obsegom, poleg tega pa se uporablja tudi padajoča polkadenca, podobna kadenci. Katera vrsta kadence nastopi, je odvisno od tega, ali gre za podredje, priredje ali naštevanje. Pravilna izbira je nemogoča brez ustrezne pomenske in stavčne razgradnje povedi.

Intonacijske segmente delimo, podobno kot tudi ločila v skladenjski rabi, na *končne* intonacijske segmente, ki so lahko kadenčni ali antikadenčni, in *nekončne* intonacijske segmente, ki so polkadenčni [SAZU 90, str. 58].

Za kadenco je značilno glasovno padanje, ki ob koncu intonacijskega segmenta doseže spodnjo mejo tonskega registra govorca. Če v intonacijskem segmentu ni posebej poudarjene besede, osnovna frekvenca znotraj takega segmenta upada, še bolj izrazito pa od zadnjega izraziteje naglašnega zloga naprej. Če pa vsebuje intonacijski segment poudarek, se osnovna frekvenca spušča od poudarjenega zloga naprej.

Za antikadenco je značilno naraščanje od zadnjega poudarjenega ali izraziteje naglašnega zloga naprej proti zgornji meji tonskega obsega govorca ali pa ustalitev na določeni višini. Polkadenca zaznamuje nekončno intonacijo, podobno kot končna intonacija je lahko rastoča in padajoča.

Končne intonacijske segmente zaznamujejo končna ločila v skladenjski rabi (pika, vprašaj, klicaj in tri pike), nekončne intonacijske segmente pa nekončna ločila v skladenjski rabi (vejica, dvopičje, podpičje ipd.). Poznamo štiri vrste stavčne intonacije.

Povedna je tista končna intonacija, ki se zaključí s kadenco. Uporablja se v povedih, izgovorjenih brez čustvene zavzetosti. Predvsem je značilna za povedi s povednim naklonom. Povedna intonacija je značilna za povedni, pogojni ali veljni glagolski naklon. To nam kažejo primeri: <Tjaša gre domov.>, <Domov pojdi.>, <Tega vam ne bi oprostila.>. Uporablja se lahko tudi za vprašalne povedi z zaimensko vprašalnico, npr. <kdo>, <kako>. V pisavi jo običajno zaznamujemo s piko na koncu povedi, pa tudi s podpičjem sredi povedi.

Vprašalna intonacija je tista, ki se končuje z antikadenco. Uporablja se v odločevalnih vprašalnih stavkih, zlasti z vprašalnico <ali>, ki je lahko tudi izpuščena, npr. <Ali si že naredil domačo nalogo?> in <Si videl nov film?>. V pisavi se zaznamuje z vprašajem.

Vzključna intonacija ima širše intonacijsko ogrodje, je pa lahko kadenčna ali antikadenčna. Uporabljam jo, kadar kaj povemo s čustveno prizadetostjo. V pisavi to nakažemo s klicajem, npr. <Domov pojdi!> in <Ali si to ti naredil?!>.

Nekončna intonacija se končuje s polkadenco. Uporabljam jo pri nekončnih segmentih. Navadno na tem mestu uporabljamo vejico. Glede na strukturo in pomen se odločimo za rastočo ali padajočo polkadenco.

5.7.3 Register

Register je v slovenski slovnici podan kot relativni tonski položaj stavčne intonacije glede na tonski obseg govorca [Toporišič 91, str. 459]. V navadnem govorjenju, za razliko od petja, navadno uporabljamo srednji pas svojega govornega tonskega območja, višje preidemo le v izjemnih primerih oz. pri poudarjanju. Povedne kadenca se navadno končajo na dnu srednjega tonskega območja.

Register se v povedi navadno zmanjšuje od leve na desno [SAZU 90, str. 58]. Tako je vsaka naslednja intonacijska enota v povedi nižja od prejšnje (zaporedje priredij). Naslednjo povedno enoto nato spet začnemo višje. Vrinjeni stavki imajo nižji register kot ostali stavki; del povedi za vrinjenim stavkom je na isti tonski višini, kot če vrinjeni stavek sploh ne bi obstajal.

5.7.4 Ritem govora

Ritem govora zaznavamo kot ubrano menjavanje med naglašeni in nenaglašeni zlogi v govoru [Toporišič 91, str. 443]. Za govora je namreč naporno, če dva ali več naglašeni zlogov izgovorimo neposredno drugega za drugim [Omerza 70, str. 119]. Ritem v govoru lahko primerjamo z enakomernim menjavanjem poudarjenih in nepoudarjenih dob v glasbi [Miksič 93]; npr. pri trodobnem taktu je poudarek vedno na prvi dobi v taktu. Le da si pri glasbi poudarki sledijo v bolj strogo omejenih časovnih razmikih kot pri spontanem govoru.

Številne raziskave po svetu so pripeljale do spoznanja, da je trajanje intervala med dvema naglasoma odvisno od števila glasov med njima in od števila zlogov v tem intervalu [Fant 96]. Tako je težko govoriti o popolni ritmičnosti proznega govora, kakršna pa je značilna za recitiranje poezije, še bolj pa je vprašljiva tako imenovana *izohroničnost govora*, ki predvideva, da je intuitivno zaznavanje naglasa neposredno korelirano z akustičnimi dogodki na govornem signalu [Cummins 96; Benguerel 86; Lehiste 77; Fant 96]. Tako pri spontanem govoru govorimo o *kvaziritmičnosti*, kar pomeni, da je lokalno povprečno trajanje intervalov med naglasi dokaj konstantno, s precejšnjim standardnim odmikom.

Čeprav merodajnih dokazov o izohronosti govora ni, se je v svetu uveljavila delitev na jezike, v katerih naj bi bili intervali med naglašeni zlogi razmeroma stalni, in na jezike, kjer so trajanja posameznih zlogov razmeroma stalna.

Angleški jezik uvrščajo v prvo skupino jezikov, francoščina pa je značilen predstavnik druge skupine jezikov. Slovenščina, hrvaščina, srbsščina in ruščina, torej jeziki s premičnim besednim naglasom, pa spodbujajo nepravilni takt [Bezljaj 71], čeprav nekateri tudi slovenščino uvrščajo v prvo skupino jezikov.

V Miksič-Kačičevi študiji [Miksič 93] so iskali korelacijo med ritmičnostjo, kot jo zaznavamo v glasbi, in ritmičnostjo pri govoru, ki jo zaznavamo pretežno intuitivno [Cummins 96]. Pri tem so predvsem opazovali trajanje notnih in zlogovnih segmentov. Zanimivo bi bilo vedeti, ali so opazili sistematičnost v trajanju intervala med naglašeni zlogi.

5.7.5 Hitrost in tempo govora

Na trajanje zlogov v tekočem govoru močno vpliva hitrost, s katero govorimo. Ljudje govorimo različno hitro, nekateri hitro, drugi zmerno hitro ali počasi [Toporišič 91, str. 460]. Priporoča se zmerna hitrost govora, tako da nas poslušalci lahko dobro razumejo tudi v posameznostih in iz njih strniti večje miselne celote. Sredstvo za uravnavanje hitrosti govora so tudi primerni premori.

Pri srednji govorni hitrosti vendar vsega ne govorimo enako hitro. Bolj pomembno izrazimo s primerno upočasnjenim izgovorom, manj važno preletimo hitreje. Jedro stavka povemo počasneje kot prehod, včasih tudi počasneje od izhodišča; vrinjeni stavek pa navadno izgovorimo hitreje od sobesedila.

Hitrost govora prilagajamo tudi naslovniku in času, ki ga imamo na voljo za razlago. Poslušanja manj vajenim govorimo počasneje, počasneje pa tudi, če razpravljamo o težjem vprašanju. Če moramo obravnavati kopico enot v kratkem času, govorimo hitreje. Hitrost govora je v skladu s čustveno obarvanostjo tematike ali pa razpoloženja pripovedujočega: resne in žalostne stvari govorimo upočasnjeno; veselo, prijetno vsebino podajamo živahno. Hitrost govora je dodatno odvisna od temperamenta govorečega. Pretirano hitrega ali počasnega govora navadno ne poslušamo najbolj sproščeno.

6. MERITVE POTEKOV PROZODIČNIH PARAMETROV ZA SLOVENSKEGA GOVORCA

V tem poglavju opisujemo meritve trajanja in osnovne frekvence, ki smo jih dodatno izvedli z namenom, da bi preverili ugotovitve, opisane v prejšnjem poglavju, in hkrati izmerili prozodične parametre slovenskega govorca [Vračar 95; Benedik 96; Rakar 95]. Besedilo poglavja je v glavnem povzeto po omenjenih diplomskih delih.

Da bi raziskali prozodične parametre v slovenskem govornem jeziku, smo morali posneti več vrst govornih posnetkov.

Za določanje inherentnega trajanja ter inherentne osnovne frekvence samoglasnikov smo posneli zbirko vnaprej izbranih logatomov. Za določanje značilnih potev intonacijske krivulje smo posneli govorca, ki je izgovarjal besedilo, vzeto iz člankov v dnevnem časopisju. Na voljo nam je bila tudi slovenska govorna baza GOPOLIS [Dobrišek 96a], ki se nanaša na področje posredovanja informacij o zračnih poletih. V bazo so vključeni začetni pozdravi, uvodne fraze, kratka uvodna poizvedovanja, daljša poizvedovanja, kratka pojasnila in potrditve ter zahvale in zaključni pozdravi. Bazo je snemalo 50 govorcev, vsak je izgovoril po 200 povedi. Za meritve tempa artikulacije in vpliva hitrosti govora na trajanje slovenskih glasov smo del baze GOPOLIS posneli pri treh različnih hitrostih govora.

Govor je bil posnet v normalnem laboratorijskem akustičnem okolju, poskušalo se je le izogibati nenavadnim zvokom, kot so pogovor v ozadju in loputanje z vrati. Posnet je bil s frekvenco vzorčenja 16 kHz in linearno kvantiziran s 16 biti na vzorec.

Govorni signal smo razčlenjevali na glasovne segmente na dva načina, samodejno in ročno. Za ročno označevanje govornih signalov smo imeli na voljo zelo učinkovit uporabniški vmesnik, ki omogoča rezanje, označevanje in predvajanje posameznih delov govornega signala [Dobrišek 95]. Pri tem lahko izbiramo med časovnim ali spektralnim prikazom govornega signala.

Za samodejno račlenjevanje govornega signala smo uporabili sistem za razpoznavanje tekočega slovenskega govora, ki temelji na prikritih Markovovih modelih [Ipšič 96].

6.1 Meritev inherentnega trajanja in inherentne osnovne frekvence za slovenske glasove

Ko poslušamo različne govorce, opazimo, da se med seboj razlikujejo tako po trajanju kot tudi po osnovni frekvenci posameznih glasov, prav tako pa so lahko

bolj ali manj usmerjeni v izgovorjavo samo enega značilnega tonemskega naglasa (akut ali cirkumfleks).

Kot primer navedimo primerjavo dveh poklicnih govorcev, ki oba govorita knjižno slovenščino [Srebot 88, str. 23–32]. Razlika v trajanju med naglašnim dolгим in nenaglašnim [a] je pri prvem govorniku 100 %, pri drugem pa 40 %. Prvi govornik je besedilo izgovarjal z nižjimi osnovnimi frekvencami kot drugi. Zato smo se odločili, da bomo inherentno trajanje in osnovno frekvenco merili le pri enem govorniku.

Za pravilno opravljene in vrednotene meritve z veliko zbirko besed je treba imeti jezikoslovno fonetično znanje in izkušnje, katerega imajo lahko le izurjeni fonetiki. Ti izberejo poklicne govorce, ki jih prej preizkusijo, ali je njihova izgovorjava primerno razločna, da so meritve dovolj merodajne. Ker si nismo mogli pomagati niti s poklicnimi govorniki niti z izšolanimi fonetiki, smo se odločili, da opravimo nekaj krajših meritev na izbranih besedah, ki jih je izgovoril govornik F. M. Izbrani govornik je predloženo besedilo izgovarjal v zborni izreki knjižne slovenščine na ljubljanski osnovi.

Inherentno trajanje in osnovno frekvenco glasov smo proučevali na zbirki 60 umetno sestavljenih nesmiselnih besed, t. i. *logatomov*, v nevtralni intonacijski legi [Vračar 95, str. 78]. Logatome smo izbrali iz obsežnega seznama nesmiselnih besed, ki so jih že predhodno v svojih raziskavah uporabili fonetiki [Srebot 88, str. 252].

Izbrane logatome sestavljata en samoglasnik in en soglasnik, ki se neprestano izmenjujeta. Soglasniki so bili izbrani tako, da so predstavljali štiri glasovne skupine: zapornike, zlitnike, pripornike in zvočnike. Samoglasniki so vedno nastopali med istimi soglasniki in v enakih strukturah zlogov. Trajanje polglasnika [@] smo izmerili v naravnih besedah, gl. [Vračar 95, str. 79].

Za vsak samoglasnik smo sestavili več logatomov, v katerih je naglašni samoglasnik nastopal v različnih strukturah zlogov. Pri oznakah strukture umetnih besed smo uporabili naslednje simbole: C soglasnik, V samoglasnik, : dolžina, ' začetek naglašnega zloga, _ podčrtani samoglasnik, ki nas zanima.

Logatomi so bili sestavljeni na naslednji način:

'V:CV	dvozložna beseda z odprtim naglašnim prvim zlogom z dolгим samoglasnikom, ki nima predhodnega soglasnika, sledi mu ponaglasni zlog
CV'CV:CV	trozložna beseda z odprtim naglašnim srednjim zlogom, ki ima predhodni soglasnik ter ima prednaglasni in ponaglasni zlog
CV'CV:C	zaprt naglašni zlog v končni legi z dolгим samoglasnikom, ki ima prednaglasni zlog
CV'CV_C	zaprt naglašni zlog v končni legi s kratkim samoglasnikom in prednaglasnim zlogom

Tabela 16: Sestava logatomov.

Prvi dve skupini predstavljata baritone, drugi dve oksitone. Vsako besedo sestavljata enak samoglasnik in soglasnik, ki se ponavljata tako dolgo, da je dosežena zelena struktura zloga oziroma besede.

Da bi se čim bolj izognili vplivu stavčne intonacije, smo logatomom levo in desno dodali niza smiselnih besed. Tako se je logatom nahajal v sredini intonacijske fraze, kjer je vpliv stavčne intonacije najmanjši. Tako smo npr. za glas [i] merili zelene vrednosti poleg ostalih tudi na naslednjih logatomih (podajamo jih s sobesedilom): reci pipíp enkrat, ne dvakrat, reci didíd enkrat, ne dvakrat, reci títiti enkrat, ne dvakrat, reci kikíki enkrat, ne dvakrat, reci sisís enkrat, ne dvakrat, reci čičič enkrat, ne dvakrat.

Na teh besedah je mogoče proučevati inherentno trajanje in osnovno frekvenco slovenskih samoglasnikov v idealnih razmerah, ker vsak samoglasnik nastopa v istih legah kot vsi ostali in je tako možno zanemariti vpliv sosednjih soglasnikov ter je mogoče primerjati samoglasnike med seboj na dva načina [Srebot 88, str. 20]: paradigmatično, skozi inherentno trajanje, in sintagmatično, v naglašeni in v nenaglašeni zlogih, obakrat v nevtralni intonacijski legi.

Vsak samoglasnik je treba proučiti glede na to, ali se nahaja v naglašeni ali v nenaglašeni zlogu; glede na zgradbo zloga: CV ali VC, kar pomeni odprt ali zaprt zlog, ter posebej v zaprti končni zlogu glede na razliko v trajanju med pričakovanim dolgim in kratkim samoglasnikom. Trajanje nenaglašeni samoglasnikov smo merili posebej v prednaglasni in v ponaglasni zlogih. Iz izmerjenih vrednosti višine osnovnega tona za akute in za cirkumflekse smo poskusili določiti vsaj približne velikosti nastopajočih skokov.

Trajanje glasov smo določili ročno, s pomočjo časovnega in spektralnega poteka govornega signala logatoma. Rezultati meritev so zbrani v tabelah 17 do 23.

6.1.1 Trajanje samoglasnikov

Trajanje naglašeni samoglasnikov je podano v prvih štirih stolpcih v tabeli 17. V naglašeni zlogih so samoglasniki najdaljši v trizložnih baritonih, kjer je naglas na odprtem zlogu, ki se začne s soglasnikom: sklop številka 2. To se ne sklada z ugotovitvami Srebotove [Srebot 88, str. 23–32], kjer so pri šolanih govornicah zaznali, da so samoglasniki najdaljši v odprtih zlogih, brez predhodnega soglasnika: sklop številka 1.

Lahko pa potrdimo ugotovitev Srebotove [Srebot 88, str. 242], da je naglasen samoglasnik, ki sodi med krajše, lahko celo krajši od kakšnega nenaglašeni samoglasnika: kratek naglasni [i] v četrtem sklopu je krajši od marsikaterega nenaglašeni samoglasnika v ponaglasni zlogu. Ta ugotovitev je pomembna, saj si na ta način lahko razložimo pojav nenaglašeni samoglasnikov, ki so daljši od njihovih naglašeni sosedov.

Iz tabele 18 je razvidno, da so samoglasniki najkrajši v prednaglasni zlogih, kar

se ujema z ugotovitvami Srebotove [Srebot 88, str. 32]. V tej legi so najkrajši zato, ker so nenaglašeni in hkrati omejeni s soglasnikom, ki jim sledi. Samoglasniki v prednaglasni legi so v povprečju za 27 % krajši od samoglasnikov v ponaglasni legi.

samoglasnik	trajanje samoglasnika v naglašenem zlogu [ms]			
	baritoni		oksitoni	
	'(C)V:CV	CV'CV:CV	CV'CV:C	CV'CV:C
i :	94	80	93	
i"				62
E:	165	146	148	
E"				82
e:	123	191	90	
a"				111
a:	144	171	227	
O"				96
o:	133	243	218	
O:	124	233	144	
u"				102
u:	108	152	104	
ɔ"				82
povprečje	127	174	146	90

Tabela 17: Trajanje naglašanih samoglasnikov v baritonih in v oksitonih.

Nadalje lahko ugotovimo, da je trajanje nenaglašanih samoglasnikov v končni legi lahko zelo dolgo, kot je to razvidno iz dveh predzadnjih stolpcev tabele 18. Ker jim ne sledi soglasnik, lahko namreč nemoteno izzvenijo.

samoglasnik	trajanje samoglasnika v nenaglašenem zlogu [ms]			
	prednaglasni zlog		ponaglasni zlog	
	CV'CV:CV	CV'CV:CV CV'CV:CV	'CV:CV	CV'CV:CV
i	36	51	52	70
E	93	86	116	126
a	78	72	113	107
O	111	80	119	112
u	78	68	123	120
povprečje	79	71	105	107

Tabela 18: Trajanje nenaglašanih samoglasnikov v prednaglasnih in v ponaglasnih zlogih.

Tabela 19 podaja primerjavo povprečnih trajanj izmerjenih naglašanih samoglasnikov in nenaglašanih samoglasnikov v prednaglasni legi; ponaglasna lega ne daje dobrih primerjav, ker so ponaglasni zlogi pogosto v izglasju.

samoglasnik	trajanje t_1 [ms]	samoglasnik	trajanje t_2 [ms]	t_1 / t_2	t_1 / t_2 [%]
i	91	i	44	2.07	48
E:	153	e	90	1.70	59
e:	135				
a:	181	a	75	2.01	41
O:	198	o	100	1.98	51
o:	167				
u:	121	u	73	1.66	60
povprečje	149		76	1.96	51

Tabela 19: Primerjava trajanja samoglasnikov v naglašeni in v prednaglasni legi.

Velja pa, da je trajanje samoglasnikov v ponaglasni legi v vseh primerih krajše od trajanja naglašanih samoglasnikov. Ta ugotovitev je še posebej zanimiva zato, ker so v Miksič-Kačičevi študiji [Miksič 93] opazili podaljšanje zadnjega zloga v besedi, pri čemer k zadnjim zlogom niso vključili le zadnjih zlogov v frazi. To trajanje naj bi celo prevladalo nad trajanjem naglašanih zlogov.

Med naglašeni samoglasniki je najdaljši [o:] s povprečnim trajanjem 198 ms, ki mu s 181 ms sledi [a:]. V študiji Srebotove [Srebot 88] je bil najdaljši samoglasnik [a:], ki mu je tesno sledil [o:]. Najkrajši je samoglasnik [i:] z 91 ms, ki mu sledi [u:] s 121 ms. Razmerje med najkrajšim [i:] in najdaljšim [o:] je 2.17 ali 46 %, kar pomeni, da traja [i:] 46 % časa trajanja [o:].

Samoglasniki v prednaglasni legi so v povprečju za 51 % krajši od naglašanih samoglasnikov.

6.1.2 Osnovna frekvenca samoglasnikov

Tabeli 20 in 21 podajata osnovno frekvenco za samoglasnike v naglašanih in v nenaglašanih zlogih. Osnovno frekvenco smo merili v sredinskem delu samoglasnika z ročnim označevanjem osnovnih period.

Rezultati meritev ne odražajo dosledno inherentnih osnovnih frekvenc, ker so pogojeni tudi z besednim naglasom, zato jih obravnavamo bolj na kratko. Prav tako bi moral izšolan fonetik presoditi, kdaj je govorec uporabil akutni in kdaj cirkumfleksni naglas. Po naši presoji je govorec logatome izgovarjal pretežno cirkumfleksirano.

Osnovna frekvenca naglašanih samoglasnikov v baritonih in v oksitonih je po-

dana v tabeli 20. Vidimo, da so povprečne vrednosti zbrane v razmiku 9 Hz okoli vrednosti 164 Hz.

samoglasnik	osnovna frekvenca samoglasnika v naglašenem zlogu [Hz]			
	baritoni		oksitoni	
	'(C)V̇:CV	CV'CV̇:CV	CV'CV̇:C	CV'CV̇C
i :	180	207	205	
i"				186
E:	139	145	143	
E"				187
e:	170	158	182	
a"				157
a:	153	129	130	
O"				126
o:	181	169	152	
O:	137	136	140	
u"				177
u:	192	187	169	
ɔ"				178
povprečje	165	162	160	169

Tabela 20: Osnovne frekvence naglašanih samoglasnikov v baritonih in oksitonih.

Ni mogoče opaziti, da bi bili samoglasniki v zaprtih zlogih višji od samoglasnikov v odprtih zlogih, kot so to v predhodni študiji ugotovili za govorca *Ju* [Srebot 88, str. 34]. *Ju* je namreč naglašene odprte zloge izgovarjal akutirano, naglašene zaprte zloge pa cirkumfektirano. Zato pa so rezultati podobni tistim za govorca *Ka* in *Pi*, pri katerih se vse povprečne vrednosti nahajajo v razmiku 4 Hz oziroma 10 Hz.

samoglasnik	osnovna frekvenca samoglasnika v nenaglašenem zlogu [Hz]			
	prednaglasni zlog		ponaglasni zlog	
	CY'CV:CV	CY'CV̇CV	'CV:CV̇	CV'CV:CV̇
		CY'CV̇:CV		
i	141	140	115	111
E	132	147	105	117
a	102	117	98	89
O	143	128	97	99
u	138	128	106	107
povprečje	131	132	114	105

Tabela 21: Osnovne frekvence nenaglašanih samoglasnikov v prednaglasnih in v ponaglasnih zlogih.

Iz tabele 21 je razvidno, da so samoglasniki v prednaglasni legi višji kot tisti v ponaglasni legi, kar se ujema z rezultati za govorca *Pi* in *Ka* [Srebot 88, str. 33–39]. Ponovno pa se to ne ujema z rezultati za govorca *Ju*, ki je ponaglašene samoglasnike izgovarjal višje kot prednaglašene, ker je odprte naglašene zloge izgovarjal akutirano. Govorec *Ju* je besede s strukturo 'CV:CV izgovarjal tako, kot je to predvideno, namreč akutirano [Rigler 68, str. 198].

Tabela 22 podaja primerjavo povprečnih vrednosti izmerjenih osnovnih frekvenc med naglašeni in med nenaglašeni samoglasniki v prednaglasni legi.

samoglasnik	f_1 [ms]	samoglasnik	f_2 [ms]	f_1 / f_2	f_1 / f_2 [%]
i:	197	i	141	1.40	72
E:	142	e	139	1.02	98
e:	170				
a:	137	a	110	1.25	80
O:	167	o	136	1.23	81
o:	138				
u:	184	u	133	1.38	72
povprečje	162		132	1.28	81

Tabela 22: Primerjava osnovnih frekvenc samoglasnikov v naglašeni in prednaglasni legi.

Med naglašeni samoglasniki je najvišji [i:] s povprečno višino 197 Hz, ki mu s 184 ms sledi [u:]. Najnižji je samoglasnik [a:] s 137 Hz, ki mu sledi [o:] s 138 Hz. Razmerje med najvišjim [i:] in najnižjim [a:] je 1,4 ali 70 %. Izkaže se, da so glasovi, ki so inherentno krajši, tudi višji ([i:] in [u:]), in obratno: glasovi, ki so inherentno daljši, so nižji ([a:] in [o:]).

Osnovna frekvenca je pri naglašeni samoglasnikih v povprečju kar za 81 % višja od osnovne frekvence pri prednaglasni samoglasnikih. Govorec je namreč logatome izgovarjal pretežno cirkumflektirano.

6.1.3 Trajanje soglasnikov

Za merjenje trajanja soglasnikov smo uporabili del slovenske govorne baze GOPOLIS, ki ga je izgovarjal naš izbrani govorec F. M. Obsegal je zbirko 266 povedi tekočega govora s področja posredovanja informacij o letalskem prometu [Dobrišek 96a]. Razčlenjevanje govora na glasovne enote je opravil programski paket za samodejno razpoznavanje tekočega slovenskega govora, ki temelji na prikritih Markovovih modelih [Ipšič 95].

Merili smo trajanje soglasnikov v VCV in CC sklopih. Tabela 23 podaja rezultate merjenj povprečnega trajanja soglasnikov. Izmerjene vrednosti so precej kraj-

še od predhodno izmerjenih vrednostmi za slovenske govorce [Bezljaj 39]. Menimo, da do tega pride, ker smo trajanje soglasnikov merili v besedah, ki smo jih vzeli iz tekočega govora, medtem so bili v Bezlajevi študiji [Bezljaj 39] soglasniki morda merjeni v ločeno izgovarjanih besedah.

soglasnik	trajanje soglasnika [ms]	
	VCV	CC
p	68	44
t	84	64
k	68	65
b	67	60
d	53	50
g	98	55
s	103	54
S	110	58
z	59	57
Z	62	58
dZ	69	66
m	62	54
n	45	34
h	98	33
l	96	50
v	59	48
r	43	45
j	42	37
f	86	85
ts	131	100
tS	84	60
w	60	31

Tabela 23: Povprečne vrednosti trajanja soglasnikov v VCV in v CC sklopih.

Iz rezultatov, zbranih v tabeli 23, ugotovimo naslednje. Soglasniki v CC sklopih so v povprečju za 23 % krajši od soglasnikov v VCV sklopih. Razlike so majhne pri zapornikih: če izvzamemo zapornik [g], so zaporniki v CC sklopih v povprečju za 15 % krajši od zapornikov v VCV sklopih. Zanimiv je soglasnik [r], ki se kot edini podaljša v CC sklopu. Soglasnik [r] je namreč težko razmejiti od polglasnika [@], ki se soglasniku [r] pridruži v sklopu CC.

6.2 Meritev trajanja slovenskih glasov pri različnih hitrostih govora

Meritve povprečnega trajanja slovenskih glasov so bile že izvedene, vendar vedno pri normalni hitrosti govora [Bezljaj 39; Srebot 88; Mihelič 93], čeprav je govorec *Pi* v študiji [Srebot 88] govoril zelo hitro. Mi smo meritve izvedli za enega moškega govorca, pri treh različnih hitrostih govora [Benedik 96; Gros 96e].

Ko iščemo podatke o vplivu hitrosti govora za trajanje glasov, jih v obstoječi literaturi ne zasledimo prav dosti. Z opazovanjem trajanja samoglasnikov pri različnih hitrostih govora sta se ukvarjala R. J. J. H. van Son in L. C. W. Pols [Son 93], vendar so njiju bolj zanimale formantne frekvence kot samo trajanje samoglasnikov. Ugotovila sta, da so samoglasniki v hitrem govoru v povprečju samo za 15 % krajši od tistih v normalnem. Menimo, da njunih rezultatov ne moremo preprosto privzeti za slovenski jezik in da jih je treba preveriti.

Treba je bilo zasnovati načrt za analizo vpliva hitrosti govora na trajanje slovenskih glasov. Znano je, da na trajanje glasov vpliva mnogo dejavnikov. Navedimo jih le nekaj: ali so glasovi naglašeni ali nenaglašeni, ali se nahajajo v odprtem zlogu ali ne, kakšen je položaj zloga v besedi, stavku in taktu, trajanje zloga ipd. Pomembno je tudi, za kateri glas gre oziroma kateri fonetični skupini pripada: je samoglasnik ali soglasnik, zvočnik ali nezvočnik ipd.

Proučevanje vpliva vseh naštetih dejavnikov na trajanje glasov ob različnih hitrostih govora je zelo obsežno, zato smo se omejili na opazovanje krajšanja in daljšanja povprečnih trajanj posameznih glasov ter poskusili poiskati zakonitosti, ki se ob spremenjeni hitrosti govorjenja odražajo v trajanju posameznih skupin glasov.

Za besedilo smo izbrali povedi iz slovenske govorne baze GOPOLIS, ki jih je izgovarjal govorec F. M. V bazi GOPOLIS je govorec besedilo izgovarjal z normalno hitrostjo. Isto besedilo je posebej v naš namen izgovoril še dvakrat: hitro (kakor hitro je mogel) in počasi (zadržano, razvlečeno). Normalno hiter govor traja 7 min 32 s, počasen govor 12 min 55 s in hiter govor 5 min 45 s.

Če smo želeli proučiti daljšanje in krajšanje slovenskih glasov pod vplivom hitrosti govora, je bilo treba najprej določiti njihovo trajanje za vsako hitrost posebej. Trajanje samoglasnikov in soglasnikov smo obravnavali ločeno. V moji doktorski disertaciji [Gros 97e, str. 95] so podani histogrami, ki prikazujejo trajanje samoglasnikov pri normalni hitrosti govora.

V tabeli 24 so zbrane vrednosti povprečnega trajanja in razmiki zaupanja za posamezne glasove in glasovne komponente pri treh hitrostih govora: normalni, počasni in hitri. Razmik zaupanja je določen za 95 % verjetnost, da se povprečna vrednost trajanja glasu nahaja v določenem intervalu.

	normalno niter govor		niter govor		pocasen govor	
	povprečno trajanje	razmik zaupanja	povprečno trajanje	razmik zaupanja	povprečno trajanje	razmik zaupanja
glas	[ms]	±[ms]	[ms]	±[ms]	[ms]	±[ms]
a	64,76	1,31	58,15	1,30	147,26	4,86
a:	131,90	2,66	100,40	3,45	232,54	9,04
E	59,28	1,42	52,35	1,82	126,79	5,72
E:	119,95	6,75	91,95	6,18	232,51	30,63
e:	98,23	4,35	79,53	2,88	205,83	9,96
@	48,26	3,56	39,58	3,79	83,96	6,61
i	56,09	1,37	49,15	1,78	130,95	5,74
i:	95,53	3,52	68,38	3,10	181,30	9,88
O	63,93	1,88	55,28	1,76	127,39	5,88
O:	140,82	8,13	87,64	9,10	239,61	24,88
o:	99,79	4,17	81,44	3,15	189,93	9,46
u	71,12	4,77	54,54	3,60	148,84	13,62
u:	128,58	9,93	103,47	3,93	266,11	29,21
odpora b	20,77	1,58	19,02	1,13	24,62	1,87
zapora b	53,13	3,76	40,28	3,30	114,02	7,63
b	74,07	3,89	59,40	3,23	138,45	7,86
odpora ts	67,27	8,79	48,58	4,08	99,15	12,51
zapora ts	50,27	7,11	51,27	8,31	127,70	12,28
ts	117,55	10,74	99,85	10,24	226,85	20,32
odpora tS	60,24	7,76	51,24	4,27	100,95	12,60
zapora tS	32,00	4,05	31,60	4,65	89,79	11,84
tS	92,24	10,31	82,83	7,03	190,74	19,40
odpora d	19,84	1,37	18,04	0,54	24,74	1,99
zapora d	49,10	2,72	37,16	2,32	117,15	7,68
d	69,01	3,46	55,27	2,47	141,95	8,26
f	78,42	8,86	60,37	9,88	165,68	22,89
odpora g	24,46	1,77	21,66	1,27	36,09	3,76
zapora g	35,96	2,18	32,83	2,55	84,29	6,90
g	57,05	2,30	52,32	2,67	115,03	7,65
h	88,39	10,92	62,18	8,85	165,00	13,41
j	50,06	4,33	32,99	2,80	115,91	10,76
odpora k	33,27	0,93	31,20	0,74	46,27	2,22
zapora k	36,69	2,32	27,12	1,81	101,32	7,25
k	69,98	2,66	58,58	2,05	146,76	8,29
l	47,88	2,14	40,27	1,72	109,83	4,45
m	66,93	3,40	59,46	3,59	138,16	7,34
m	46,32	2,39	39,73	1,87	106,41	5,51
odpora p	22,11	1,12	21,58	0,86	24,43	2,83

zapora p	65,50	3,63	50,09	3,03	157,90	9,42
p	87,71	3,62	71,77	2,87	182,42	9,56
r	50,46	1,67	45,46	1,43	94,01	3,34
s	89,46	5,29	72,75	4,20	177,78	8,73
S	84,03	4,73	70,03	3,50	199,81	12,26
odpora t	23,17	0,92	24,35	0,69	27,28	1,57
zapora t	64,72	5,27	49,34	1,92	161,16	8,36
t	87,94	5,29	73,81	1,93	188,61	8,34
v	50,35	2,94	46,31	2,70	104,06	6,59
z	70,12	2,56	60,42	2,32	138,92	6,69
Z	75,00	10,31	61,21	6,97	133,07	17,34
w	54,64	4,97	42,53	3,70	114,95	10,21
zv. zapora	47,23	1,89	37,08	1,59	108,55	4,74
nezv. zapora	53,16	2,49	41,54	1,47	135,21	4,90

Tabela 24: Povprečna vrednost trajanja in razmik zaupanja glasov in glasovnih komponent pri treh hitrostih govora. Razmik zaupanja je določen za 95 % verjetnost, da se povprečna vrednost nahaja v danem intervalu.

6.2.1 Analiza parov

Ker je govorec v vseh treh primerih izgovarjal isto besedilo, vsakokrat z različno hitrostjo, smo imeli priložnost, da opravimo statistično obdelavo časov trajanj v besedilu istoležnih glasov s postopkom, ki temelji na analizi parov [Pavešič 81, str. 173].

Postopki statistične analize para naključnih vzorcev temeljijo na treh predpostavkah: naključno izbrana vzorca sta neodvisna, populaciji sta normalno porazdeljeni in varianci obeh populacij sta enaki.

Statistično obdelavo naključno razporejamo nad pari poskusnih enot in ne med vsemi enotami. Tako naključnost razporeda omejimo s tem, da en člen para obdelamo z eno, drug člen pa z drugo obdelavo. Oceno razlike učinkov obdelav dobimo iz vsakega para posebej. Tako smo iz variance izločili razpršenost med pari. S primerjavami parov lahko dosežemo večjo natančnost le, če smo pare sestavili tako, da so odkloni med enotami iz različnih parov večji kot odkloni med enotami v paru.

Izračunali smo nepristranske cenilke za povprečno vrednost in varianco razlik med trajanji istoležnih parov glasov [Gros 97e, str. 99]. Pare smo sestavljali tako, da smo v par jemali trajanja istoležnih glasov: prvič pri počasnem in normalno hitrem govoru; drugič pri normalno hitrem in hitrem govoru.

glas	počasen : normalen govor		normalen : hiter govor	
	povprečna vrednost [ms]	razmik zaupanja ±[ms]	povprečna vrednost [ms]	razmik zaupanja ±[ms]
a	82,50	4,44	6,60	1,17
a:	100,64	8,68	31,50	3,29
E	67,51	5,26	6,93	1,55
E:	112,56	26,08	28,00	7,32
e:	107,60	7,99	18,71	2,84
@	35,70	5,99	8,68	3,71
i	74,86	5,48	6,94	1,61
i:	85,77	8,67	27,16	3,55
O	63,46	5,25	8,64	1,74
O:	98,79	25,18	53,18	12,30
o:	90,14	8,24	18,35	2,91
u	77,72	11,50	16,58	4,53
u:	137,53	24,19	25,11	10,96
odpora b	3,85	1,59	1,75	1,02
zapora b	60,89	6,46	12,86	3,27
b	64,38	6,69	14,67	3,58
odpora ts	31,88	8,86	18,70	8,60
zapora ts	77,42	11,06	-1,00	7,85
ts	109,30	14,13	17,70	3,97
odpora tS	40,71	8,47	9,00	5,02
zapora tS	57,79	10,10	0,40	3,70
tS	98,50	13,87	9,40	5,58
odpora d	4,89	1,92	1,81	1,27
zapora d	68,05	6,72	11,94	2,42
d	72,95	7,10	13,74	2,77
f	87,26	18,51	18,05	9,56
odpora g	11,63	3,18	2,80	1,34
zapora g	48,32	6,16	3,13	2,84
g	57,97	6,84	4,73	2,65
h	76,61	15,33	26,21	11,36
j	65,85	8,36	17,07	3,03
odpora k	13,00	2,21	2,07	0,95
zapora k	64,62	6,47	9,57	1,75
k	76,78	7,27	11,39	1,79
l	61,95	4,14	7,61	2,10
m	71,23	6,57	7,47	2,83
m	60,10	4,48	6,59	2,02
odpora p	2,32	2,98	0,53	1,22

zapora p	92,41	8,50	15,41	3,48
p	94,71	8,73	15,94	3,65
r	43,55	3,02	5,00	1,80
s	88,32	6,02	16,70	3,50
S	115,78	9,67	14,00	3,66
odpora t	4,11	1,69	-1,19	0,99
zapora t	96,44	7,15	15,39	4,96
t	100,67	7,07	14,13	4,95
v	53,71	6,78	4,04	2,93
z	68,80	6,11	9,70	2,21
Z	58,07	18,99	13,79	13,23
w	60,31	8,94	12,11	3,84
zveneča zapora	61,32	4,02	10,15	1,67
nezveneča zapora	82,05	3,96	11,62	2,12

Tabela 25: Analiza parov. Ocena povprečne vrednosti razlike v trajanju in razmika zaupanja za istoležne pare pri različnih hitrostih govora. Razmik zaupanja je določen za 95% verjetnost, da se povprečna vrednost nahaja v danem intervalu.

Rezultati, podani v tabeli 25, v dobršni meri potrjujejo trditev, da se trajanje dolgih glasov spreminja bolj kot trajanje kratkih glasov. Trajanje zapor zlitnikov in zvenečih pripornikov se spreminja zelo malo. Pri samoglasnikih se ozka glasova krajšata manj kot ostali dolgi, polglasnik pa celo bolj kot nekateri kratki. Pri hitrem govoru se močno krajša drsnik [j], ponekod celo izgine, npr. iz zloga [Eja] nastane [Ea] in iz sklopa glasov [nj] le nekoliko mehkejši [n']. Razlike med krajšanjem in daljšanjem posameznih glasov nas napeljujejo na misel, da ima vsak glas svojo spodnjo mejo trajanja, pod katero ne sme iti, in ta meja nastopa za različne glasove pri različnih hitrostih govora različno.

6.2.2 Normirano daljšanje in krajšanje glasov

Podatke, uporabljene v analizi parov, smo predhodno normirali na trajanje glasu oziroma glasovne komponente za normalno hiter govor [Gros 97e, str. 103]. Nato smo podatke, ki smo jih uporabili pri analizi parov, za vsak par posebej normirali glede na trajanje glasu oz. glasovne komponente za normalno hiter govor in šele nato nad njimi izvedli analizo parov, gl. tabelo 26.

glas	normirana razlika: počasen : normalen govor		normirana razlika: normalen : hiter govor	
	povprečna vrednost	± razmik zaupanja	povprečna vrednost	± razmik zaupanja
a	1,31	0,07	0,09	0,02
a:	0,78	0,07	0,24	0,02
E	1,16	0,09	0,11	0,03
E:	0,89	0,17	0,22	0,05
e:	1,14	0,10	0,17	0,03
@	0,81	0,15	0,15	0,08
i	1,39	0,10	0,11	0,03
i:	0,92	0,09	0,27	0,03
O	1,02	0,08	0,12	0,03
O:	0,73	0,18	0,36	0,07
o:	0,98	0,12	0,17	0,03
u	1,14	0,17	0,20	0,05
u:	1,07	0,18	0,17	0,07
odpora b	0,23	0,07	0,05	0,03
zapora b	1,32	0,18	0,20	0,06
b	0,92	0,10	0,17	0,04
odpora ts	0,50	0,13	0,23	0,07
zapora ts	1,92	0,54	-0,14	0,30
ts	0,96	0,12	0,15	0,03
odpora tS	0,73	0,13	0,10	0,05
zapora tS	1,97	0,43	-0,04	0,14
tS	1,13	0,17	0,07	0,05
odpora d	0,29	0,09	0,03	0,03
zapora d	1,48	0,15	0,20	0,05
d	1,10	0,10	0,17	0,04
f	1,15	0,23	0,21	0,11
odpora g	0,52	0,12	0,07	0,05
zapora g	1,42	0,20	0,04	0,09
g	1,03	0,12	0,07	0,05
h	1,14	0,36	0,23	0,12
j	1,40	0,17	0,25	0,05
odpora k	0,42	0,06	0,04	0,02
zapora k	2,02	0,24	0,20	0,05
k	1,13	0,11	0,15	0,02
l	1,53	0,15	0,08	0,05
m	1,18	0,13	0,10	0,04
m	1,50	0,12	0,07	0,04
odpora p	0,17	0,18	-0,02	0,05

zapora p	1,54	0,18	0,21	0,04
p	1,12	0,11	0,16	0,03
r	0,94	0,08	0,06	0,03
s	1,06	0,08	0,17	0,03
S	1,42	0,13	0,15	0,04
odpora t	0,23	0,08	-0,09	0,03
zapora t	1,67	0,13	0,17	0,03
t	1,23	0,09	0,12	0,02
v	1,24	0,21	0,03	0,06
z	1,03	0,10	0,12	0,03
Z	0,86	0,33	0,13	0,16
w	1,27	0,22	0,19	0,06
zveneča zapora	1,42	0,10	0,16	0,04
nezveneča zapora	1,79	0,10	0,16	0,03

Tabela 26: Normirana povprečna razlika trajanj in razmik zaupanja za istoležne glasovne pare. Razmik zaupanja je določen za 95% verjetnost, da se povprečna vrednost nahaja v danem intervalu.

Najmanj, skoraj nič, se krajšajo oziroma daljšajo odpore zapornikov. Pri odporah zapornikov se zopet najbolj spreminjata [k] in [g].

Zanimivo se obnašajo zapore zlitnikov, ki se močno daljšajo in prav nič ne krajšajo. Nasprotno velja za trajanje odpore zlitnika skupaj s pripadajočim pripornikom. Ta se močno krajša in razmeroma malo daljša. Za pripornike velja, da se nezveneči bolj spreminjajo kot zveneči. Zvočniki se bolj daljšajo in manj krajšajo. Temu splošnemu pravilu se ne pokoravata le oba drsnika. Pri samoglasnikih se dolgi glasovi bolj daljšajo in manj krajšajo; obratno velja za kratke samoglasnike.

Glasovi, ki so sami po sebi kratki, z izjemo odpor zapornikov, se bolj daljšajo in manj krajšajo in se pri hitrem govoru hitro bližajo svojemu minimalnemu trajanju. Zlitniki in dolgi samoglasniki se manj podaljšujejo, ko se približajo zgornji meji trajanja.

Kaže, da za vsak glas obstaja spodnja meja trajanja, pri kateri je glas še razpoznaven in pod katero pri izgovarjavi večinoma ne gremo, ter zgornja meja trajanja, katere pri neprisiljenem govorjenju ne prekoračimo.

6.3 Meritev tempa artikulacije pri različnih hitrostih govora

Tempo artikulacije, ki ga podaja število zlogov, izgovorjenih v časovni enoti, je pomemben dejavnik pri določanju prozodičnih parametrov pri sintezi govora. Če želimo govor sintetizatorja upočasniti ali pohitrili, je treba poznati tempo

artikulacije za različne hitrosti govora. Tovrstne meritve za slovenski jezik še niso bile izvedene, zato smo se odločili za preprosto analizo trajanja zlogov na posnetkih besedila, izgovarjanega z različnimi hitrostmi.

Tempo artikulacije se v govoru neprestano spreminja. Na trajanje zloga in posredno tudi na tempo artikulacije, vpliva mnogo dejavnikov [Grosjean 75; Zeller 94]: število zlogov v besedi, naglašenost zloga, hitrost govora, mesto in tip naglasa v besedi, položaj zloga v frazi oziroma povedi in taktu in število glasov v zlogu.

V našem primeru smo od navedenih dejavnikov upoštevali le hitrost govora, število zlogov v naglasni celoti in položaj naglasne celote v stavku, čeprav se zavedamo, da smo s tem zaobšli enega od najbolj vplivnih dejavnikov, podatek o naglasu v besedi.

Dolžino izgovorne celote smo definirali kot število zlogov v eni naglasni celoti: beseda vključno z breznaglasnicami, ki se nanjo naslanjajo. Tako izgovorno celoto smo poimenovali *takt* [Bezljaj 71].

Taktovne skupine se tvorijo po dveh načelih. Tako imenovane izdišne ali ekspiratorne celote so podrejene fiziološkimi procesom dihanja. Nanje vplivajo kapaciteta pljuč, tempo, poudarek, jakost govora itd. Logične, sintaktične taktovne skupine ali sintagme pa se oblikujejo po psihičnem členjenju govora. Omejene so s premori. Jezikovna kultura stremi za tem, da se fiziološke izdišne skupine taktov krijejo z logičnimi.

Tempo artikulacije smo merili za štiri položaje izgovornih celot v stavku: osamljene takte, takte za premorom, takte pred premorom in sredinske takte ali takte med premori.

Za preizkuse smo imeli na voljo govorno bazo moškega govorca v obsegu trikrat po 172 povedi. Govorec je povedi izgovoril trikrat: normalno hitro, počasi in hitro. Grobo razčlenjevanje govornega signala na glasovne komponente je opravil programski paket za samodejno razpoznavanje tekočega slovenskega govora [Ipšič 96]. Dobljene oznake smo ročno popravili.

Besedilo smo razčlenili v takte. Posneto govorno bazo smo tudi poslušali, da smo preverili, da so bile vse breznaglasnice zares izgovorjene brez naglasa, saj se v nasprotnem primeru struktura takta spremeni. Določiti smo morali tudi položaj premorov v govorjenem besedilu. Nismo razlikovali med daljšimi oziroma krajšimi premori.

Tabela 27 podaja trajanje zlogov pri normalni hitrosti govora za štiri položaje izgovornih celot v povedi: za osamljene takte, takte za premorom, takte pred premorom in takte med premori. Podobni tabeli za počasen in hiter govor se nahajata v moji doktorski disertaciji [Gros 97e, str. 106].

Tempo artikulacije se v govoru neprestano spreminja, tako da je vsaka določitev tempa artikulacije nujno samo približna oz. povprečna. Zaradi tega pri merjenju trajanja enega segmenta govora nismo nikoli povsem prepričani, kolikšna je vloga

vseh ostalih dejavnikov, ki vplivajo na trajanje tega segmenta. To je prav gotovo eden od pomembnih vzrokov precejšnje razpršenosti rezultatov naših meritev.

število zlogov	1	2	3	4	5	6	7	8
izolirani takti	3.03	4.27	5.44	6.02	7.76			
takti za premorom	4.26	6.33	7.11	7.18	8.76			
takti med premori	5.42	5.80	6.41	6.78	7.20	6.73	7.16	7.25
takti pred premorom	3.19	4.47	5.05	5.65	5.94	6.69	6.03	6.11

Tabela 27: Tempo artikulacije pri normalni hitrosti govora.

Ugotovimo, da tempo artikulacije narašča z večanjem števila zlogov v taktu. To pomeni, da na splošno daljše besede izgovarjamo hitreje kot krajše. S tem se pri razumevanju govora olajša razmejevanje med besedami in poveča razumljivost. Tako smo potrdili ugotovitev, ki jo podaja Riedi [Riedi 95], da besede z različnim številom zlogov težijo k temu, da bi trajale enako dolgo.

Tempo artikulacije neposredno za premori je navadno večji od povprečnega, pred premori pa manjši. Na vrednost tempa artikulacije v zadnjem taktu pred premorom najbolj vpliva daljšanje zadnjega zloga.

Grafi so podobni za vse tri hitrosti govora [Gros 97e, str. 107,108]. V tekočem govoru se res najhitreje izgovarjajo besede za premorom, počasneje tiste, ki niso omejene s premori, in najpočasneje tiste, ki so pred premorom. Izjema je tu le počasen govor, kjer je tempo artikulacije med premori za spoznanje višji od tistega za premorom. Zanimivo je, da se osamljeni takti ne izgovarjajo vedno najpočasneje. Enozložni takt, ki ni omejen s premori, je izgovarjan hitreje kot enozložni takt za premorom.

Pokazano je bilo, da je možno odvisnost tempa artikulacije od števila zlogov v besedi približno izraziti s pomočjo matematičnega modela, logaritemske funkcije [Bakran 84]. To se izkaže za prikladno predvsem v primerih, ko govorna baza vsebuje premalo taktov z določenim številom zlogov.

Trajanje takih taktov ocenimo z matematičnim modelom. Variacije tempa artikulacije aproksimiramo z logaritemsko krivuljo: $TA = b + m \cdot \ln n$, kjer je TA tempo artikulacije, n število zlogov v besedi, b in m pa koeficienta, ki ju moramo določiti.

Parametra b in m smo določili s pomočjo optimizacije. Pri optimizaciji nismo upoštevali taktov s številom zlogov n , večjim od 5. Taki takti so namreč razmeroma redki in zaradi tega je tudi napaka pri določanju tempa artikulacije pri teh taktih precej velika. Število n smo ponekod celo zmanjšali na 4, saj se je zgodilo, da v besedilu ni bilo takta s petimi zlogi ali pa je bilo število vzorcev premajhno (le eden ali dva), da bi izmerjeni vrednosti tempa artikulacije lahko zaupali.

Za začetni vrednosti b in m pred optimizacijo smo izbrali vrednosti, ki jih lahko določimo iz podatkov za izmerjene vrednosti tempa artikulacije za takte z enim

in dvema zlogoma. Nato smo s spreminjanjem obeh parametrov poiskali minimum kriterijske funkcije in določili vrednosti za parametra b in m za vse tri hitrosti govora, kot so prikazani v tabeli 28.

položaj takta	normalno hitro		počasi		hitro	
	b	m	b	m	b	m
izolirani takti	2.95	2.19	1.38	0.83	5.16	1.47
takti za premorom	4.47	2.20	1.81	1.24	6.30	2.09
takti med premori	5.27	1.11	2.18	0.96	6.64	0.91
takti pred premorom	3.22	1.72	1.54	0.95	3.90	2.00

Tabela 28: Koeficienta b in m matematičnega modela za različne hitrosti govora in lege takta.

Ugotovimo, da se izmerjena in izračunana krivulja lepo ujemata za število zlogov, manjše od 5, potem pa se razideta. Do tega pride predvsem zaradi majhnega števila vzorcev. Od števila 5 zlogov naprej se nekatere krivulje obnašajo povsem nepredvidljivo, vendar kažejo določeno težnjo, da bi ostale na istem nivoju ali da bi celo padale. Menimo, da do tega pride zato, ker so takti nad 5 zlogov že predolgi, da bi se tempo artikulacije lahko še kar naprej povečeval; možno pa je tudi, da pride do podobnega pojava kot v glasbi, kjer dobijo večdobni takti še enega ali več pomožnih poudarkov.

6.4 Merjenje značilnih intonacijskih krivulj za potek stavčne intonacije

Da bi določili značilne poteke intonacijskih krivulj, smo opazovali poteke osnovne frekvence na zbirki 50 testnih povedi. Povedi so bili vzete iz dnevnega časopisja. Govorec A. R. jih je izgovarjal z normalno hitrostjo govora. Sprva smo morali za vse signale določiti poteke osnovne frekvence, ki smo jih nato ročno stilizirali.

Za samodejno določanje osnovne frekvence iz govornega signala je bilo razvitih mnogo algoritmov [Nöth 90, str. 96]. Odločili smo se, da uporabimo algoritem AMDF (angl. Average Magnitude Difference Function) [Ross 74], ker je preprost, časovno nezahteven in deluje v časovnem prostoru govornega signala [Gros 91; Rakar 95].

Funkcija povprečne razlike amplitud AMDF je variacija avtokorelacijske funkcije. Namesto korelacije vhodnega signala pri različnih zakasnitvah opazujemo absolutno vrednost razlike med vhodnim signalom in zakasnjanim vhodnim signalom. Prednost takega pristopa je zlasti v tem, da algoritem ne zahteva množenja, kar znatno pospeši izvajanje na nekaterih vrstah procesorjev.

AMDF izračunamo za vsako vrednost zakasnitve preko okna dolžine L vzorcev.

Poleg te funkcije uporabimo tudi odločitveno logiko, ki vsebuje predhodna znanja o zvenečih in nezvенеčih glasovih in pravih nastopanja le-teh [Rakar 95, str. 28–30]. S tem je določanje osnovne frekvence bolj točno in zanesljivo.

Dobljene krivulje poteka višine osnovnega tona smo ročno linearizirali in tudi odstranili vpliv mikroperturbacij v osnovni frekvenci in vpliv oslabiljenega tonemskega naglasa [Rakar 95, str. 40]. Ker je bil postopek ročen, je bila končna linearizirana krivulja močno odvisna od osebe, ki je postopek izvajala. Predvsem je vprašljivo, kako natančno je bil izločen vpliv besednega naglasa.

Pri iskanju značilnih intonacijskih potekov smo se naslonili na predhodne ugotovitve slovenskih jezikoslovcev [Toporišič 69]. Njihove ugotovitve smo potrdili z opazovanjem krivulj osnovne frekvence na testnih povedih. Dobljeni rezultati so le okvirne narave in jih je treba dodatno preveriti na večjem številu povedi. Prav tako je treba natančno definirati kriterije, ki jih uporabimo pri linearizaciji in stilizaciji krivulje osnovne frekvence.

7. NASTAVLJANJE PROZODIČNIH PARAMETROV

Predvsem na naravnost sintetiziranega govora bistveno vplivata bogatost intonacijskih in ritmičnih vzorcev [Collier 90]. K nadsegmentnim prozodičnim lastnostim sodi še časovni razvoj amplitude glasov. Ker pa je glasnost govornega segmenta razmeroma stalna skozi vse njegove pojave v govoru, ob predpostavki, da besedilo izgovarjamo neprizadeto, jo bomo obravnavali kot lastnost, ki je že vsebovana v govornem segmentu. Naglase v besedah pa bomo oblikovali s primernim podaljšanjem trajanja in ustreznim potekom osnovne frekvenca.

V tem poglavju govorimo o določanju trajanja in osnovne frekvenca na podlagi fonetičnega prepisa besedila z označenimi mesti naglasa. Poglavje je razdeljeno v tri ločene dele.

Prvi del govori o predpripravi fonetičnega zapisa besedila. Ta del pokriva: zlogovanje besed, oblikovanje taktov in členitev besedila na fraze in določanje tonemskega naglasa.

Drugi del govori o modeliranju trajanja. Zajema nastavljanje trajanja glasov in zlogov ter nastavljanje trajanja premorov.

Tretji del se nanaša na modeliranje osnovne frekvenca. Ločeno obravnava nastavljanje tonemskega besednega naglasa in oblikovanje stavčne intonacije.

7.1 Predpriprava fonetičnega zapisa besedila

Za uspešno nastavljanje parametrov trajanja in osnovne frekvenca je potrebna predpriprava fonetičnega prepisa besedila z naslednjimi koraki: deljenje besed na zloge, določanje tipa besede glede na naglasno mesto, oblikovanje taktov, grupiranje besedila oz. členitev besedila na fraze, določanje položaja takta in določanje vrste tonemskega naglasa.

Samodejno deljenje besed na zloge je precej težavna naloga, čeprav se zdi na prvi pogled preprosta. Pravila za deljenje slovenskih besed so opisana v slovenski slovnici [Toporišič 91, str. 86]. Avtor, podobno kot Srebotova [Srebot 75a; Srebot 92] ugotavlja, da se soglasniki ob samoglasnikih zbirajo v zlog po načelu nezvočnik-zvočnik-samoglasnik-zvočnik-nezvočnik. Pri tem je lahko nezvočnikov oz. zvočnikov na isti strani samoglasnika tudi po več.

Pri določanju zlogov upoštevamo osnovno načelo, da se mora v zlogu nahajati vsaj en samoglasnik, ki predstavlja jedro zloga. Le v primeru dvoglasnika ali diftonga lahko zlog vsebuje dva samoglasnika. Pri določanju zbirke pravil za deljenje besed na zloge smo upoštevali napotke iz literature [Toporišič 91, str. 86; Weilguny 93], nekaj pravil smo dodali po lastni presoji [Gros 97e, str. 116, 117].

Dobljeni zlogi niso v vseh primerih pravilni. Nepravilnost se pojavi pri določanju mej med zlogi, ker se mejni soglasnik lahko pojavi v nepravilnem zlogu. Število zlogov pa je vselej pravilno, kar je pomembno pri nastavljanju tempa artikulacije. Lega naglašene zloga je prav tako pravilna.

Besede delimo glede na položaj naglašene zloga v besedi na baritone in na oksitone. Bariton je večzložna beseda, ki nima naglasa na zadnjem zlogu. Oksiton je beseda z naglasom na zadnjem ali edinem zlogu. Baritoni in oksitoni se razlikujejo po trajanju zlogov in po poteku krivulje osnovne frekvence pri tonemskem naglaševanju.

Takt v okviru tega dela definiramo kot besedo, ki ji pridružimo vse brez naglasnice, ki se na besedo naslanjajo glede naglasa.

Breznaglasnice so nenaglašene besede, ki se v govorni verigi naslanjajo na naglašene besede, zato se imenujejo naslonke ali klitike. V slovenščini poznamo dve vrsti naslonk [Toporišič 91, str. 57]: *enklitike* – naslonke, ki se naslanjajo na naglašeno besedo pred seboj, in *predslonke* – naslonke, ki se naslanjajo na naglašeno besedo za seboj.

Za slovenske brez naglasnice je značilno, da se vedejo v glavnem kot predslonke [Orešnik 83]. Naslonke, ki se nahajajo med dvema naglašema besedama, pripišemo prejšnji besedi (gre za enklitiko) le v primeru, da naslonka ni predlog, veznik ali členek, kajti te naslonke so vedno predslonke. Če naslonke izgovarjamo osamljene, dobijo stavčni naglas oziroma poudarek.

Pomenske enote v povedi so ločene z daljšimi ali krajšimi prekinitvami, ki jih v pisavi pogosto zaznamujejo ločila [Omerza 70, str. 116]. Tako je govor razdeljen na različne segmente, ki nastopajo kot posledica različnih stavčnofonetičnih pojavov [Toporišič 91, str. 437], imenujemo jih *fraze*.

Besedilo smo samodejno členili na fraze, pri tem smo pod frazo pojmovali segment govora, obdan s premori, ko govorimo o členitvi besedila s premori, intonacijski segment, ko govorimo o nastavljanju stavčne intonacije, in ritmični segment, ko govorimo o ritmičnem oblikovanju govora.

Za pravilno členitev besedila na fraze bi bilo treba sprva opraviti natančno skladenjsko in pomensko razgradnjo besedila [Toporišič 91, str. 442]. Ob menjavanju tematike besedila potrebujemo celo pragmatično znanje o povedanem, kajti takrat se premori dodatno podaljšajo. Sintetizator slovenskega govora v trenutnem stanju razvoja take razgradnje še ne omogoča, zato smo členitev besedila na fraze močno poenostavili.

Besedilo samodejno členimo na fraze na naslednjih mestih:

- na koncu povedi: ta mesta so v pisavi zaznamovana s končnimi ločili, kot so pika, vprašaj, klicaj in tri pike v skladenjski rabi ipd.;
- na meji med stavki iste povedi, zlasti v priredju: pred prirednimi vezniki, zlasti protivnimi in posledičnimi;
- med vrivkom, pristavkom, dostavkom, izpostavkom, pastavkom in preosta-

lim delom povedi ali stavka: ta mesta so v pisavi navadno zaznamovana z nekončnimi ločili, kot so vejica, dvopičje, podpičje, vezaj ipd.;

- po spremnem stavku premege govora, če je pred dobesednim navedkom: to mesto je v pisavi zaznamovano z vejico in z narekovajem;
- v podredu pred prilastkovim, časovnim in predmetnim odvisnikom: ta mesta so v pisavi navadno zaznamovana z vejico pred podrednim veznikom.

Poznavanje položaja takta v prozodičnem segmentu je pomembno pri nastavljanju trajanja zlogov ob upoštevanju tempa artikulacije. Ločimo štiri položaje takta: takt pred premorom, takt za premorom, sredinski takt in izolirani takt.

Pravilna določitev tonemskega naglasa je zelo pomembna pri zagotavljanju naravnosti sintetiziranega govora. Iz navajane raziskave [Srebot 88, str. 179–181] sledi, da je naglas odvisen predvsem od govorca. Človek se tonemskega naglasa priuči že v zgodnji mladosti in ga skozi življenje ne spreminja [SAZU 90, str. 75]. Tako nekateri govorijo izrazito cirkumfektirano, drugi akutirano. Obstajajo govorci, pri katerih celo izkušeni jezikoslovci težko vrsto tonemskega naglasa. Lega besede oz. takta v intonacijskem segmentu dodatno vpliva na tonemski naglas, in sicer: začetna lega pospešuje akut, končna s padajočim jedrom pa cirkumfleks. Končna lega z rastočim jedrom vsiljuje akut [Srebot 88, str. 245]. Analiza tonemskih naglasov pri treh govorcih, *Ju*, *Ka* in *Pi* v študiji Srebotove [Srebot 88, str. 179–181], pokaže, da je bil pri upoštevanju naglasa najbolj sistematičen govorec *Ju*. *Ju* namreč izgovarja odprte naglašene zloge logatomov akutirano, zaprte pa cirkumfektirano. Pri daljših besedah pa prevladuje akut.

Ker nismo imeli na voljo slovarja vseh besed z označeno vrsto tonemskega naglasa, smo se odločili, da tonemski naglas določamo na naslednji način. Začetni takt v intonacijskem segmentu je akutiran. Končni takt v intonacijskem segmentu je akutiran, če ima rastoče jedro (polkadenca in antikadenca), in je cirkumfektiran, če ima padajoče jedro (kadenca). Če je naglašeni zlog odprt, je takt izgovorjen akutirano. Če je naglašeni zlog zaprt in ima takt tri ali manj zlogov, je izgovorjen cirkumfletirano. Če je naglašeni zlog zaprt in ima takt več kot tri zloge, je izgovorjen akutirano.

Tako določanje naglasa se je po dolgotrajnem slušnem preizkušanju sintetiziranega govora pokazalo kot najbolj primerno. Umetni govor, ki je vseboval samo akute ali samo cirkumflekse, je bil slišati preveč enoličen. Cirkumfleksni naglas pri večzložnih baritonih, kjer je število zlogov preseгло tri, je bil slišati nenaraven.

Izbrani mešani naglas je bil najbolj všeč tudi poslušalcem, ki so se udeležili preizkusa sintetizatorja govora. Poslušalcem smo isto besedilo predvajali trikrat. V prvem primeru smo vsem besedam nastavili akutni naglas, v drugem primeru cirkumfleksnega. V tretjem primeru smo oba tipa naglasa menjavali na način, ki je opisan zgoraj. Poslušalce smo prosili, da se odločijo, katera različica sintetiziranega govora jim je najbolj všeč. Kar 81 odstotkov poslušalcev se je

odločilo za mešani naglas, 5 odstotkov za cirkumfleksnega, 14 odstotkov poslušalcev pa ni zaznalo nobene razlike med predvajanimi različicami sintetiziranega govora.

7.2 Modeliranje trajanja glasov in zlogov

Sintetizirani govor, pri katerem so upoštewane le srednje vrednosti trajanja posameznih glasov, zveni močno nenaravno [Epitropakis 93a]. Zato je treba modeliranju trajanja glasov in premorov posvetiti posebno pozornost.

Postopek modeliranja trajanja smo razdelili na dva dela: nastavljanje trajanja govornih enot in nastavljanje trajanja premorov.

Na trajanje glasov v določenem glasovnem in prozodičnem okolju vpliva mnogo različnih dejavnikov, med njimi so najbolj pomembni [Epitropakis 93a]: glasovni kontekst, v katerem se glas nahaja (predhodni in naslednji glas); segmentna identiteta (pove ime glasu, ki mu določamo trajanje); naglašenost zloga (ali se glas nahaja v naglašenem, prednaglasnem ali ponaglasnem zlogu); dolžina besede, ki ji glas pripada (število zlogov v besedi); položaj glasu znotraj besede (zadnji, predzadnji, predpredzadnji zlog); poudarjenost besede (stopnja poudarjenosti besede po jakostni členitvi besedila); položaj besede znotraj fraze in dolžina fraze (začetna, sredinska, končna, izolirana lega); hitrost govora (zmerna, hitra, počasna ali vmesne stopnje).

V preteklosti je bilo predlaganih kar nekaj pristopov, ki so poskušali zajeti vpliv čim večjega števila naštetih dejavnikov na trajanje glasov oziroma zlogov [Klatt 87; Epitropakis 93a; Möbius 96].

Ne glede na to, kakšen pristop je bil uporabljen pri modeliranju trajanja govornih enot in kakšen je bil učni postopek, vključujejo opisani sistemi podobno informacijo. Tako so pri mnogih jezikih prišli do podobnih spoznanj:

- nekateri glasovi so inherentno daljši od ostalih [Allen 87; Srebot 88, str. 242; Crystal 90; van Santen 93], tako so samoglasniki v povprečju daljši od soglasnikov;
- zadnji zlog na koncu fraze ali pred premorom se podaljša [Allen 87; Vogel 96; Fant 96];
- besede, ki imajo le skladenjsko vlogo, se izgovarjajo hitreje kot tiste, ki so pomembne tudi z vsebinskega gledišča [Sproat 95];
- glasovi v naglašanih zlogih so navadno daljši od tistih v nenaglašanih [Allen 87; Srebot 88; Tzoukerman 95], tako samoglasniki kot soglasniki;
- če je samoglasnik prvi glas v besedi, je daljši, kot v primeru, če bi bil pred njim še kak soglasnik [Tzoukerman 95; Srebot 88, str. 242].

Velikosti govornih enot, ki jim s pomočjo omenjenih postopkov prirejamo traja-

nje, so različne: glas (v literaturi ga pogosto imenujejo kar segment), zlog in beseda.

Mnogi pristopi k modeliranju trajanja govornih enot predpostavljajo, da lahko pojav trajanja zajamemo v *govornem segmentu ali glasu*. Zadnja dognanja pa so pokazala, da je trajanje večjih govornih enot – zlogov manj spremenljivo kot trajanje govornih enot, ki so manjše od zloga [Barbosa 93; Breen 95]; tako *zlog*, in skozi njega tudi takt, predstavlja bolj zanesljivo enoto za določanje trajanja govornih enot. Metoda napovedovanja trajanja segmentov, če poznamo trajanje zloga, mora biti v tem primeru dovolj kompleksna [Campbell 91].

Kot osnovna enota pri modeliranju trajanja lahko služi tudi *beseda*. Trajanje besede in njenih segmentov je odvisno od njene intrinzične in ekstrinzične dolžine [Ferreira 93]. Za vsako besedo ima govorec v svojem spominu shranjeno inherentno ali *intrinzično trajanje*. To trajanje se v kontekstu z drugimi besedami prilagodi *ekstrinzičnemu trajanju*, ki ga določata hitrost govora in položaj besede znotraj fraze.

Pri določanju trajanja segmentov govora smo ocenili, da *takt* kot skupek zlogov besede, skupaj z breznaglasnicami, predstavlja boljše govorno enoto za napovedovanje trajanja akustičnega signala kot kompleksno in slabo definirano določanje trajanja za posamezne glasove. Trajanje takta predvidimo kot funkcijo železne hitrosti govora, števila zlogov v besedi oziroma v taktu in položaja takta in zloga v frazi. Postopek napovedovanja trajanja glasov, ko poznamo trajanje takta, mora biti dovolj zanesljiv.

Dejavniki in parametri, ki jih upoštevamo pri napovedovanju trajanja glasov in taktov, so skladenjske, fonetične in fiziološke narave.

Fonološki dejavniki: inherentno fonološko trajanje (vrsta glasu: samoglasnik ali soglasnik; glasovni kontekst, v katerem se glas nahaja: npr. sklop VCV ali sklop CC pri soglasnikih; naglašenost zloga: ali se glas nahaja v naglašenem ali nenaaglašenem zlogu; položaj zloga v taktu: naglašeni, prednaglasni, ponaglasni, vmesni, začetni, končni; položaj glasu znotraj besede: npr. če se beseda začne s samoglasnikom, se le-ta podaljša; zgradba zloga: npr. ali se glas nahaja v odprtem ali v zaprtem zlogu).

Fiziološki dejavniki: minimalno trajanje posameznega glasu. Skladenjski dejavniki: položaj takta znotraj fraze (takt pred premorom, takt za premorom, zadnji takt v frazi, izolirani takt). Velikost takta: določena je s številom zlogov v taktu, tako poznamo enozložne, dvožložne takte ipd. Hitrost govora: vpliva na trajanje takta, v katerem se glas nahaja.

Vpliv naštetih dejavnikov na trajanje glasovnega segmenta ocenimo na podlagi fonološkega in jezikovnega opisa besedila.

Za nastavljanje trajanja govornih enot za slovenski jezik smo uporabili *dvostopenjski pristop* napovedovanja trajanja. Sprva dva medsebojno neodvisna modela ločeno predlagata intrinzično trajanje govornih enot na nivoju glasov in ekstrinzično trajanje govornih enot na nivoju taktov.

Sprva model za *nastavljanje intrinzičnega trajanja glasov* govornim enotam v trajanju enega takta priredi inherentno ali *intrinzično trajanje*, ki ga dobimo kot vsoto intrinzičnih trajanj glasov, vsebovanih v taktu. Model pri nastavitvi trajanja glasov upošteva dejavnike, ki izhajajo iz segmentnih značilnosti govora, načina artikulacije in položaja glasov ter kontekstualnih vplivov. Pod segmentnimi značilnostmi govora v tem kontekstu pojmujeemo nižjenivojske značilnosti govora, ki delujejo nad govornimi enotami na nivoju glasov in glasovnih sklopov. Sledi model za *nastavljanje ekstrinzičnega trajanja taktov*. Ko se takti vključujejo v večje govorne enote (frazе, povedi), se *skrajšujejo* ali *podaljšujejo* v skladu z zahtevami višjenivojskih prozodičnih pojavov in privzamejo *ekstrinzično trajanje*. Pri določanju ekstrinzičnega trajanja takta upoštevamo različne parametre, kot so položaj takta v frazi, izbrana hitrost govora in dolžina takta (število zlogov v taktu). Model pri določanju trajanja takta upošteva dejavnike, ki obravnavajo nadsegmentne značilnosti govora in ki izhajajo iz vrste in položaja takta v frazi ter iz želene hitrosti govora.

Za preklop med obema nivojema za razliko od Ferreire [Ferreira 93], ki za preklopno govorno enoto jemlje besedo, za preklopno govorno enoto izberemo *takt*. S tem se izognemo nevarnosti, da bi breznaglasnice, ki so pogosto kratke, enozložne besede, obravnavali kot samostojne enozložne besede, katerim pripada razmeroma nizek tempo artikulacije.

Sledi postopek usklajevanja oz. prilagajanja intrinzičnega trajanja takta na ekstrinzično trajanje. To dosežemo s podaljševanjem oz. skrajševanjem intrinzičnega trajanja takta. Pri tem je treba upoštevati, da se vsi glasovi znotraj takta ne skrajšujejo oz. podaljšujejo v enaki meri. Iz predvidenega trajanja zloga, besede ali takta izpeljemo trajanje glasu ali pa trajanje glasu popravimo glede na znani tempo artikulacije izgovorne celote.

Za pravilno upoštevanje rezultatov modeliranja trajanja različnih govornih enot iz obeh modelov smo definirali postopek usklajevanja s translacijo, ki intrinzično trajanje takta, določeno z vsoto intrinzičnih trajanja posameznih glasov v taktu, prilagodi zahtevanemu ekstrinzičnemu trajanju in vzpostavi pravilna razmerja med trajanjem posameznih glasov znotraj takta. Pri tem je treba pravilno upoštevati, da se trajanje glasov ob spremembi trajanja celotnega takta ne spreminja v enaki meri [Gros 97e, str. 131–136; Gros 99].

Modela za določanje intrinzičnega trajanja glasov in ekstrinzičnega trajanja taktov temeljita na rezultatih, ki smo jih dobili z meritvami trajanja glasov v izbranih ločeno izgovarjanih besedah in tudi tekočem govoru za prvi model, in na rezultatih meritev trajanja taktov na tekočem branem govoru za drugi model. Postopek prilagajanja skupnega trajanja glasov v taktu zahtevanemu trajanju takta temelji na rezultatih, ki smo jih dobili pri merjenju spreminjanja trajanja posameznih glasov ob različnih hitrostih govora.

7.2.1 Trajanje samoglasnikov

Najprej nastavljamo trajanje samoglasnikov. Na trajanje samoglasnikov vpliva mnogo dejavnikov, med katerimi so najbolj pomembni: naglašenost ali nenaglašenost zloga, v katerem se samoglasnik nahaja; odprtost in zaprtost zloga, v katerem se samoglasnik nahaja; lega nenaglašene samoglasnika v besedi (ponaglasni ali prednaglasni zlog ipd.); ali se samoglasnik nahaja v končnem zlogu ali ne (trajanje samoglasnika v končnem zlogu pred premorom se podaljša, še zlasti takrat, ko je končni zlog odprt [Srebot 88, str. 242]); ali se samoglasnik nahaja na začetku besede in sam tvori zlog (tedaj se trajanje samoglasnika podaljša [Srebot 88, str. 242]); tonemski naglas (naglašeni samoglasniki s cirkumfleksnim naglasom so krajši od naglašeni samoglasnikov z akutnim naglasom [Srebot 88, str. 242]).

Pri nastavljanju trajanja samoglasnikov smo upoštevali prvih pet dejavnikov. Vpliv tonemskega naglasa na trajanje naglašeni samoglasnikov smo posredno zajeli že s tem, ko smo zaprtim zlogom priredili cirkumfleksni naglas, odprtim pa akutnega.

Nastavljanje trajanja samoglasnikov poteka v dveh korakih. Najprej izvedemo nastavljanje začetnih vrednosti. Pri naglašeni samoglasniku v primeru baritona vzamemo za začetno vrednost trajanje samoglasnika iz prvega stolpca ali drugega stolpca v tabeli 17, glede na to, ali je bariton naglašen na predzadnjem ali na predpredzadnjem zlogu; v primeru oksitona vzamemo za začetno vrednost trajanje samoglasnika iz tretjega stolpca ali četrtega stolpca v tabeli 17. Če samoglasnik ni naglašen, obstajata dve možnosti. Samoglasnik se nahaja v prednaglasnem zlogu: za začetno vrednost vzamemo trajanje samoglasnika v prvem ali drugem stolpcu tabele 18, glede na to, ali je zlog odprt ali zaprt. Samoglasnik se nahaja v ponaglasnem zlogu: za začetno vrednost vzamemo trajanje samoglasnika v tretjem ali četrtem stolpcu tabele 18, glede na to, ali je beseda dvozložna ali ne.

V drugem koraku izvedemo popraviljanje začetnih vrednosti. Trajanje samoglasnika v končnem zlogu pred premorom podaljšamo: za 10 % v zaprtem zlogu in za 20 % v odprtem zlogu. Do vrednosti za daljšanje samoglasnika v končnem zlogu smo prišli s slušnimi preizkusi različnih nastavitvev.

7.2.2 Trajanje soglasnikov

Na trajanje soglasnikov vpliva mnogo dejavnikov, med njimi so najbolj pomembni: vrsta zloga, v katerem se soglasnik nahaja (naglašen ali nenaglašen); položaj soglasnika v zlogu (trajanje prvega soglasnika v zlogu je daljše kot trajanje zadnjega soglasnika v zlogu); število soglasnikov v zlogu (kopičenje večjega števila soglasnikov v zlogu vpliva na njihovo krajše trajanje; ali se soglasnik

nahaja v končnem zlogu ali ne (trajanje soglasnika v končnem zlogu pred preomorom se podaljša, predvsem trajanje zvočnika, ki lahko izzveni).

Od navedenih dejavnikov smo upoštevali prvega, tretjega in četrtega. Meritve trajanja soglasnika v soglasniških sklopih je treba za slovenski jezik še opraviti, vsaj na splošno za primer zaporniškega, priporniškega in mešanega sklopa, saj razdelitev trajanja soglasnikov v sklope CC in VCV ne daje popolnoma naravnih nastavitvev trajanja za vse možne položaje soglasnikov.

Trajanje soglasnika nastavljam v dveh korakih. Sprva nastavimo začetne vrednosti. Preverimo, ali se soglasnik nahaja v CC ali v VCV sklopu; glede na to mu priredimo ustrezno trajanje iz prvega ali iz drugega stolpca v tabeli 23.

Sledi popravljanje začetnih vrednosti. Preverimo, ali se soglasnik nahaja v naglašnem ali v nenaglašnem zlogu. Če se soglasnik nahaja v nenaglašnem zlogu, pustimo njegovo trajanje nespremenjeno. Če se soglasnik nahaja v naglašnem zlogu, mu trajanje podaljšamo za 40 %, v skladu z Bezlajevimi ugotovitvami [Bezlaj 39]. Če se soglasnik nahaja v končnem zlogu, mu podaljšamo trajanje za 10 %. Do te vrednosti smo prišli s slušnimi preizkusi različnih nastavitvev.

7.2.3 Ekstrinzično trajanje taktov

Ritem govora v jezikih, kot je angleški, določa razmeroma enakomerno menjavanje naglašnih in nenaglašnih zlogov. Ker so v slovenskem jeziku vse besede z izjemo breznaglasnic naglašene, že same po sebi s svojimi naglasi prispevajo k slušni zaznavi ritma.

Da bi v sintetičnem govoru vsaj delno ustvarili vtis kvaziritmičnega menjavanja naglašnih in nenaglašnih zlogov, smo besedam oz. taktom z večjim številom zlogov priredili krajše trajanje. To smo dosegli s spreminjanjem tempa artikulacije taktov glede na število zlogov v taktu in glede na položaj takta v frazi.

Tempo artikulacije dobimo kot razmerje med številom izgovorjenih zlogov in časom artikulacije [Shaughnessy 95], kjer so izključeni vsi premori ne glede na njihov značaj ali trajanje.

Tempo artikulacije za posamezen takt je odvisen od naslednjih dejavnikov [Natvig 94]: števila glasov v taktu, stopnje poudarjenosti takta, besedne vrste besed, ki takt sestavljajo, velikosti takta (število zlogov v taktu, položaja takta v frazi ali oz. ritmični skupini (začetna, končna, sredinska in izolirana ali osamljena lega)) in od hitrosti govora (normalno hitra, počasna, hitra in morebitne vmesne stopnje).

Pri določanju ekstrinzičnega trajanja takta upoštevamo zadnje tri dejavnike na naslednji način. Glede na izbrano hitrost govora in položaj takta v frazi iz tabele 28 izberemo ustreznega parametra b in m . Izmerimo število zlogov v besedi in s pomočjo enačbe za določanje tempa artikulacije izračunamo tempo artikulacije

za takt. Ekstrinzično trajanje takta dobimo kot vsoto trajanj posameznih zlogov v taktu.

Ker je vpliv naglašenosti zloga upoštevan že pri nastavljanju intrinzičnega trajanja, ga ne upoštevamo pri nastavljanju ekstrinzičnega trajanja. Tega tudi sicer nismo predvideli, ko smo merili trajanje zlogov pri različnih hitrostih govora.

7.2.4 Preverjanje uspešnosti napovedovanja trajanja

Pri preverjanju uspešnosti pravilnega napovedovanja trajanja smo ponovno uporabili povedi iz baze GOPOLIS, ki jih je govorec bral s tremi hitrostmi govora. Meje med glasovi so bile označene ročno.

Sprva smo ocenili, kako dosledno v smislu hitrosti govora je govorec izgovarjal besedilo, nato smo primerjali trajanje glasov pri naravnem govoru in trajanje glasov, napovedano z dvonivojskim pristopom, z uporabo različnih postopkov usklajevanja intrinzičnega trajanja z ekstrinzičnim.

Če želimo preveriti pravilnost napovedovanja trajanja glasov, moramo vedeti, kako se trajanje glasov v enakem kontekstu ohranja pri istem govorniku v tekočem govoru. Za oceno smo izbrali niz besed <Ob kateri uri>, ker se je ta niz pogosto pojavljal v bazi in ker se vedno pojavlja na začetku intonacijskega segmenta. Tako lahko pričakujemo, da bo v večini primerov izgovorjen z enako hitrostjo. Ugotovili smo, da je govorec govoril z razmeroma stalno hitrostjo, saj se posamezna trajanja glasov pri različnih izgovorjavah istega besedila le malo razlikujejo [Gros 97e, str. 137]. Pri normalni hitrosti govora dobimo standardno deviacijo 8,2 ms in povprečni absolutni odklon od srednje vrednosti 5,3 ms. Ti dve vrednosti sta zelo nizki pri hitrem govoru in večji pri počasnem.

Besedilo iz baze GOPOLIS, ki ga je govorec izgovoril s tremi hitrostmi govora, smo podali sintetizatorju govora, ki je za vsak glas napovedal njegovo trajanje, v skladu z dvostopenjskim modelom napovedovanja trajanja glasov. Pri tem smo uporabili tri različne postopke usklajevanja ekstrinzičnega trajanja z intrinzičnim: proporcionalnega, skaliranje in translacijo. Dobljene vrednosti ekstrinzičnega trajanja glasov smo primerjali s trajanjem istoležnih glasov v naravnem govoru.

Obseg testne baze izgovorjav je podan v tabeli 29:

hitrost govora	število povedi	število besed	število glasov
normalno hiter govor	172	1400	5433
hiter govor	49	607	2351
počasen govor	60	800	2900

Tabela 29: Opis testne baze izgovorjav pri vrednotenju pravilnosti napovedovanja trajanja govornih segmentov.

Rezultati, ki smo jih dobili pri normalni hitrosti govora, so nadvse obetavni [Gros 97e, str. 138]. Povprečna vrednost absolutne razlike med trajanjem glasov pri naravnem govoru in med trajanjem glasov, napovedanim z metodo translacije, znaša le 11 ms, kar je le za 6 ms več kot povprečno odstopanje trajanja istoležnih glasov od povprečne vrednosti pri naravnem govoru. Če slednji vrednosti prištejemo 2 ms, za kolikor privzamemo, da znaša napaka pri ročnem označevanju glasovnih segmentov, se omenjena razlika zmanjša na 4 ms.

Še boljše ujemanje naravnega in napovedanega trajanja glasov ugotovimo pri naglašeni samoglasnikih, kjer povprečna vrednost absolutne razlike med trajanjem glasov pri naravnem govoru in med trajanjem glasov, napovedanim z metodo translacije, znaša le 8 ms.

Standardna deviacija napovedanega trajanja od naravnega trajanja glasov znaša 15.4 ms, za naglašene samoglasnike le 13.2 ms, kar je zelo podobno rezultatom pravilnosti napovedovanja trajanja v znanem sistemu MITalk, ki še vedno velja za vzor natančnosti modeliranja trajanja glasov, kjer so dosegli standardno deviacijo 17 ms [Allen 87, str. 98]. Po podatkih iz navedene literature [Huggins 71; Klatt 76; Allen 87] in [Sorin 87], je tako odstopanje pod mejo najmanjše zaznavne spremembe trajanja (JND ali just noticeable difference) v tekočem govoru.

Menimo, da pri hitrem in pri počasnem govoru intrinzično trajanje glasu pri normalni hitrosti govora ni dovolj merodajno kot izhodiščno trajanje glasu. Predvidevamo, da bi se v primeru pravilno izbranih razmerij v trajanju glasov znotraj takta (intrinzična trajanja glasov) napovedano ekstrinzično trajanje glasu, napovedano z metodo translacije, približalo naravnemu trajanju glasov tudi pri počasnem in pri hitrem govoru.

Ustreznost napovedanega trajanja glasov smo preverili tudi s slušnimi testi na poslušalcih. Predvajali smo jim dve različici sintetiziranega govora. Prvič smo trajanje glasov napovedali samodejno z uporabo usklajevanja s translacijo. V drugem primeru smo sintetizatorju govora posredovali trajanje glasov iz naravnega govora. Poslušalce smo prosili, naj ocenijo, katera različica jim bolj ugaja. V več kot polovici primerov so poslušalci izjavili, da ne zaznajo nobene razlike med obema različicama sintetiziranega govora, nekajkrat so celo menili, da je sintetični govor s samodejno določenim trajanjem glasov bolj prijeten kot govor, kjer je trajanje vzeto iz naravnega govora.

Na koncu še izpostavimo, da imajo vrednosti, ki jih uporabljamo v modelih pri napovedovanju trajanja, fizikalen pomen in niso abstraktne uteži, kot npr. pri nevronskih mrežah, zato je možno njihovo fino nastavljanje, ker si lahko predstavljamo, na kaj vplivamo z njihovim spreminjanjem.

Menimo, da pri določanju začetnega približka trajanja glasu brez upoštevanja strukture zloga, strukture besede in lege posameznih glasov ter njihovih inherentnih trajanj ne bi bilo mogoče z naknadnimi popravki določiti prave vrednosti razmerij med njimi.

7.3 Nastavljanje trajanja premorov

Pri nastavljanju trajanja premorov smo za izhodišče vzeli vrednosti trajanja premorov na različnih mestih členitve besedila, kot jih je izmeril Toporišič [Toporišič 91, str. 443]. Te vrednosti smo spreminjali v glavnem v okviru predlaganih intervalov in s slušnimi preizkusi nastavili končne vrednosti za posamezna trajanja.

mesto členitve	Trajanje premora [ms]			
	po slovenski slovnici	zmerno hiter govor	počasen govor	hiter govor
?	800 do 860	860	1032	258
.	700	700	840	210
!	580 do 640	640	768	192
:	580 do 640	640	768	192
;	580 do 640	640	768	192
...	580 do 640	900	950	300
–	420	350	420	105
,	420	350	420	110
pred pred veznikom <zato>	270 do 350	250	300	70
pred veznikoma <in>, <ter>	30 do 70	100	130	50
med ostalimi deli priredja	180 do 240	240	310	100
pred podrednim stavkom	180 do 240	190	220	90
konec paragrafa	–	1400	1600	700

Tabela 30: Trajanje premorov na različnih mestih členitve v besedilu. Drugi stolpec podaja vrednosti po Toporišiču [Toporišič 91, str. 443] preostali trije stolpci podajajo vrednosti trajanja premorov, kot smo jih določili s slušnimi preizkusi.

V hitrem govoru se meje med besedami zabrišejo, prav tako se večina premorov močno skrajša ali pa povsem izostanejo. Vrednosti trajanja premorov za hitri in za počasni govor smo nastavljali izključno s slušnimi preizkusi, saj tovrstne meritve za slovenski govor še niso bile opravljene. Končne vrednosti trajanja premorov so zbrane v tabeli 30.

7.4 Modeliranje osnovne frekvence

Čeprav je vprašljivo, koliko Slovencev dela razliko pri izgovarjavi homonimnih leksemov, pri katerih je tonemskost pomenskorazločevalna [Šuštaršič 96, str. 42], je tonemski naglas nepogrešljiv za pravilno zborna izreko slovenskega besedila. Z željo, da bo sintetiziran govor slišati čim manj enoličen, smo se odločili, da besedni naglas oblikujemo v tonemski obliki.

Zaradi narave oblikovanja poteka osnovne frekvence v slovenskem jeziku smo

pri nastavljanju osnovne frekvence sintetiziranega govora uvedli *dvostopenjski princip*, podoben Shaughnessyjevem postopku [Shaughnessy 77], in razdelili oblikovanje poteka osnovne frekvence na dva dela: oblikovanje besednega naglasa in oblikovanje stavčne intonacije.

V nadaljevanju podpoglavja opisujemo oba postopka, pred tem pa opisujemo pristope k modeliranju osnovne frekvence za ostale jezike.

Pri samodejnem določanju poteka osnovne frekvence je treba najprej oblikovati prozodične fraze, šele nato se oblikuje časovni potek osnovne frekvence. V svetu se je uveljavilo več formalizmov za zapis ali transkripcijo časovnega poteka krivulje osnovne frekvence.

Čeprav potek prozodičnih parametrov ni nujno povezan s skladenjsko strukturo govornjene povedi [t'Hart 90], večina sistemov za sintezo govora predpostavlja, da skladnja močno vpliva na potek krivulje osnovne frekvence govornjenega, tako na oblikovanje prozodičnih fraz kot tudi na oblikovanje naglasa. Intonacijske fraze namreč pogosto sovpadajo s skladenjskimi frazami [Toporišič 91, str. 437; Dutoit 93, str. 34].

Nekateri sistemi, predvsem samodejni sistemi za dialog z uporabnikom, pri napovedovanju prozodičnih parametrov upoštevajo tudi semantične in pragmatične lastnosti govornjenega [Sorin 94; Haumont 94].

Za samodejno napovedovanje prozodičnih skupin ter vzorca poteka krivulje osnovne frekvence se uporabljajo klasični pristopi skladenjske razgradnje povedi ter novejši stohastični modeli. V prvo skupino sodijo *skladenjski razčlenjevalniki*, katerih delovanje je podrobno opisano v navedeni literaturi [O'Shaughnessy 89; Bachenko 90]. Razčlenjevalnik na funkcijske grupe [Lieberman 92] so uporabili v [Karn 96]; razčlenjevalnik ATN (Augmented Transition Network) je uporabil Mumolo [Mumolo 96], razčlenjevalnik DCG (Definite Clause Grammar) pa Traber [Traber 93].

V drugo skupino sodijo *stohastični modeli*. Pri njih gre za samodejno klasifikacijo in napovedovanje tipa prozodične fraze s prikritimi Markovovimi modeli [Batliner 93; Chou 96] in z odločitvenimi drevesi [Riley 94] z upoštevanjem skladenjskih kategorij besed.

V svetu so razvili več sistemov za transkripcijo poteka krivulje osnovne frekvence [Dutoit 93, str. 33]:

- *Tonetiko* ali *britansko šolo* je prvi predlagal Crystal [Crystal 69] in z njo opisal intonacijo britanske angleščine. Govor je razdelil na tonske skupine, ki jih sestavljajo zlogi. Zloge nadalje razvršča na poudarjene in nepoudarjene (dinamika), naglašene (kombinacija poudarka in tonemskega naglasa) ter na zloge, ki pripadajo izhodišču, glavi, jedru ali repu tonske skupine. Jedrnim zlogom pripisuje jedrni ton, ki je lahko padajoč, naraščajoč, padajoč-naraščajoč in naraščajoč-padajoč.

Ker se smer poteka osnovne frekvence nadaljuje iz jedra na rep, Halliday

- jedro in rep združi [Halliday 70]. O'Connor in Arnold uporabljata v svojem modelu intonacije sedem jedrnih tonov [O'Connor 73].
- *Pri stilizacijskih postopkih* je krivulja osnovne frekvence predstavljena z odseki, dobljenimi z interpolacijo med točkami osnovne frekvence. Poznamo več vrst stilizacijskih postopkov: kvadratične krivulje [Hirst 83; Hirst 93], povezavo elementov naraščanja in padanja – sistem RFC (angl. rise-fall-connection) [Taylor 92; Taylor 93] in linearno interpolacijo. Ta način stilizacije krivulje osnovne frekvence zadnjih trideset let razvijajo v okviru mednarodne fonetične organizacije IPO (International Phonetic Association) [IPA 89] in na Institutu v Eindhovnu [t'Hart 90]. Za mnoge evropske jezike so bile odsekoma linearne transkripcije poteka osnovne frekvence določene ročno [Cohen 67; t'Hart 73]. Uporabljajo se tudi samodejni postopki [Scheffer 88; d'Alessandro 95; Bosch 95; Grigoriu 96].
 - *Pierrehumbertov formalizem* ali *ameriška šola* [Pierrehumbert 88], ki se je kasneje razvil v:
 - *sistem ToBI* (TOne and Break Indices) [Silverman 92]. Povedi so razdeljene na intonacijske fraze, ki se nadalje delijo na vmesne fraze, te pa sestavljajo zlogi. Osnovna tonska nivoja sta visoki (H-high) in nizki (L-low). Definirana sta za lokalni frekvenčni obseg in ne izražata absolutnih vrednostih osnovne frekvence. Tonska nivoja sta definirana na koncu vsake intonacijske ter vsake vmesne fraze ter na vsakem naglašenem zlogu, skupaj s številnimi diakritičnimi znamenji (-, +, %, *, !), ki določajo njihovo interpretacijo. Sprva je bil ToBI formalizem uporabljen za opis intonacije ameriške angleščine, v zadnjem času pa ga preizkušajo tudi za japonščino JToBI [Pierrehumbert 88, Campbell 96] in za nekatere evropske jezike – nizozemščino [Gusenhoven 84], francoščino [Marten 87], madžarščino in romunščino [Ladd 83], italijanščino [Grice 92], nemščino GToBI [Grice 96], srbsščino in hrvaščino [Inkelas 88] in švedščino [Bruce 77]. Na ICSLP'96 je bil predstavljen postopek za samodejno tvorjenje JToBI zapisa [Campbell 96] za japonski jezik.
 - *Sistem INTSINT*, (*INTERNational Transcription System for INTonation*) je nadgradnja stilizacijske metode [Llisteri 94; Hirst 95] in omogoča opis poteka osnovne frekvence tako na globalnem nivoju – intonacija, kot tudi na lokalnem nivoju – mikroprozodične spremembe v osnovni frekvenci. V krivulji osnovne frekvence se po posebnih predpisih poiščejo značilne točke. Transkripcija poteka osnovne frekvence je podana z interpolacijo med značilnimi točkami, pri tem sta pomembna njihov položaj glede na ortografski zapis besedila in njihova tonska višina. Opis slovenske stavčne intonacije v sistemu INTSINT je podan pri Dobnikarju [Dobnikar 96];
 - *SAMPROSA*, (SAM PROsodic Alphabet) je nadgradnja SAMPA zapisa (SAM Phonetic Alphabet) za prozodične oznake [Wells 92]. Razvita je bila v okviru evropskega projekta SAM (Speech Input/Output Assessment Methods).

Predstavlja računalniško berljivi seznam prozodičnih oznak, ki izvirajo iz različnih prej omenjenih sistemov za transkripcijo poteka osnovne frekvence.

Pri samodejnem oblikovanju osnovne frekvence besedila se uporabljajo naslednji postopki, ki predpostavljajo, da je vhodno besedilo že razdeljeno na prozodične fraze:

- *nevronske mreže* [Chen 92; Traber 93; Sonntag 97];
- *stohastični postopki*, ki temeljijo na prikritih Markovovih modelih ali odločitvenih drevesih [Sagisaka 92; Epitropakis 93; Hirschberg 93; Yamashita 94; López-Gonzalo 96; Huang 97]; za učenje stohastičnih modelov uporabljajo zelo obsežne govorne zbirke z ročno označenimi poteki prozodičnih parametrov;
- pristopi na podlagi *postopkov za transkripcijo poteka krivulje osnovne frekvence*: generiranje potekov krivulje osnovne frekvence iz oznak ToBI z uporabo linearne regresije [Black 96] in uporaba modela Arnolda in O'Connorja [Jensen 93];
- *superpozicijski modeli*: potek krivulje osnovne frekvence modelirajo kot vsoto lokalne in globalne komponente [Chen 92; Yamashita 94], najbolj znan je Fujisakijev superpozicijski model [Fujisaki 93], ki so ga uporabili za angleški jezik [Fujisaki 95; Hirai 95], za japonsščino [Fujisaki 96b; Nakai 96; Katae 96], grščino [Galanis 96b], nemščino [Portele 94], kitajščino [Ni 96], švedščino [Ljungqvist 93; Horne 96] in slovenščino [Dobnikar 96].
Fujisakijev superpozicijski model je bil uporabljen tudi v sistemu za spreminjanje prozodičnih parametrov [Minematsu 96].

7.4.1 Oblikovanje besednega naglasa

Besedni naglas oblikujemo tako, da vsaki naglašeni besedi oz. taktu predpišemo enega od obeh značilnih tonemskih naglasov in nato oblikujemo potek osnovne frekvence za to besedo oz. takt. Oblikovanje besednega naglasa poteka v petih korakih: nastavljanje začetnih vrednosti osnovnih frekvenc, nastavljanje skoka pri tonemskem naglasu, omejitev skokov, interpolacija in razgibavanje osnovnih frekvenc.

Ker so samoglasniki pglavitni nosilci tona v govorjenem besedilu, smo menili, da moramo nastavljanju osnovne frekvence samoglasnikov posvetiti največ pozornosti. Osnovne frekvence samoglasnikov smo obravnavali kot ogrodje za potek osnovne frekvence preostalih zvenceh glasov v taktu. Sprva smo določili osnovne frekvence le njim in prvemu ter zadnjemu zvencemu soglasniku v taktu.

Pri nastavljanju začetnih vrednosti samoglasnikov smo, v skladu z ugotovitvami Srebotove [Srebot 88, str. 33–39], upoštevali naslednje dejavnike: inherentno višino samoglasnika, položaj zloga, v katerem se samoglasnik nahaja (začetni,

končni, sredinski), vrsto zloga (odprt ali zaprt) in naglašenost zloga (naglašen, ponaglasni, prednaglasni). Začetne vrednosti za osnovne frekvence samoglasnikov smo odčitali iz tabel 17 in 18, ki smo ju dobili z meritvami pri našem govorcu. Pri nastavljanju začetnih vrednosti ne nastavimo začetne osnovne frekvence le samoglasnikom, temveč tudi prvemu in zadnjemu zvenečemu soglasniku v taktu. To je treba zaradi končne interpolacije vrednosti osnovnih frekvenc, kjer ti dve točki potrebujemo za izračun vmesnih točk. Ti dve vrednosti smo nastavili s slušnimi preizkusi.

Dodatno smo upoštevali tudi položaj takta v frazi. Ker je prva beseda v frazi pogosto poudarjena, smo prvemu taktu v frazi nastavili nekoliko višje začetne vrednosti.

V slovenščini poznamo dva tonemska naglasa, to sta akut in cirkumfleks. Sprva taktu priredimo vrsto naglasa. Nato za vsak takt oblikujemo potek osnovne frekvence po enem izmed *značilnih vzorcev osnovne frekvence*. Pri tem upoštevamo: vrsto naglasa (akut ali cirkumfleks), število zlogov v taktu in položaj naglašene zloga (oksiton ali bariton).

Vsakemu taktu priredimo enega izmed petih značilnih vzorcev osnovne frekvence, ki so grafično predstavljeni na sliki 32 v navedeni literaturi [Gros 97e, str. 147]: bariton akut, oksiton akut, dvozložni bariton cirkumfleks, trozložni bariton cirkumfleks in oksiton cirkumfleks.

Če gre za akutni naglas, ločimo dva značilna vzorca osnovne frekvence. Prvi je značilen za baritone, drugi za oksitone. Pri cirkumfleksnem naglasu nastavljamo tri vzorce osnovne frekvence. Prvi je značilen za oksitone, druga dva za baritone. Tako razlikujemo med dvozložnimi in trozložnimi baritoni. Sprva smo bariton cirkumfleksu nastavili le en značilni vzorec osnovne frekvence. Vendar je bil sintetizirani govor močno enoličen, ker so se isti naglasni vzorci neprestano ponavljali. Zato smo se odločili, da dvozložnim in trozložnim bariton cirkumfleksom priredimo različna vzorca osnovne frekvence. Vzorec za trozložne bariton cirkumflekse se že močno približa akutiranim oblikam vzorca osnovne frekvence, le da je tonsko naraščanje bolj strmo.

Vse vzorce osnovne frekvence smo določili na podlagi poslušanja besed, katerim smo želeli prirediti določeno vrsto tonemskega naglasa. Prav tako smo za posamezni naglas omejili višine skokov v osnovni frekvenci, ki jih dopustimo med sosednjimi samoglasniki.

Pri nekaterih kombinacijah glasov se lahko pripeti, da je skok navzgor oziroma navzdol prevelik in je pri poslušanju moteč. Do tega pride predvsem pri določenih zaporedjih samoglasnikov, ko gre za skok iz inherentno nizkega samoglasnika, npr. [a], na inherentno visoki samoglasnik, npr. na [i], ko je le-ta naglašen (cirkumfleks) ali ponaglašen (akut). Da bi odpravili to pomanjkljivost, smo uvedli dodatno omejevanje skokov, kjer te, če so preveliki, zmanjšamo na predpisano vrednost, ki smo jo ponovno določili s slušnimi preizkusi.

Nazadnje določimo osnovne frekvence še preostalim zvenečim soglasnikom. Ker

gre za dopolnjevanje poteka osnovne frekvence v želeno obliko in višino, jim nastavimo vmesne vrednosti z interpolacijo med vrednostmi osnovne frekvence v ogrodju. Pri krajših besedah nenastavljenih osnovnih frekvenc soglasnikov ni več in se ta korak preskoči.

Pri določanju vmesnih vrednosti osnovne frekvence smo preizkusili tri vrste interpolacije: interpolacijo po metodi BSpline [Newman 84, str. 315–325], interpolacijo z Bezierovimi krivuljami in linearno interpolacijo.

V zadnjem modulu sistema uporabimo postopek sinteze govornega signala TD-PSOLA, ki temelji na združevanju difonskih govornih enot. Postopek v osnovni obliki, ki smo jo uporabili, za vsak zvoneč glas zahteva le eno vrednost osnovne frekvence, ki ustreza sredini trajanja glasu. Zato ne preseneča, da smo ugotovili, da vsi trije postopki interpolacije vrednosti osnovnih frekvenc določijo frekvenčne poteke govornega signala, ki se med seboj slušno ne razlikujejo. Prvi dve metodi celo popravljata tudi vrednosti osnovnih frekvenc v ogrodju, ki smo jih že predhodno natančno nastavili. Zato smo se odločili, da pri določanju osnovnih frekvenc preostalih zvonečih soglasnikov uporabimo linearno interpolacijo, ki je hitra in preprosta.

Nazadnje osnovne frekvence razgibamo s prištevanjem naključnih vrednosti iz manjšega frekvenčnega obsega, podobno kot to predlaga C. Sorin [Sorin 89]. Tudi na ta način želimo doseči manjšo monotonost sintetiziranega govora.

Podatke o prozodičnih parametrih besedila izpišemo v obliki niza, v katerem so sledijo glasovi in ločila z oznakami za trajanje ter oznakami za meje med takti (#).

Glas je opisan s tremi podatki: oznaka SAMPA glasu, trajanje [ms], osnovna frekvenca [HZ].

Ločilo je opisano le s trajanjem: oznaka SAMPA ločila, trajanje [ms], 00.

Podajamo primer nastavitve prozodičnih parametrov za vhodno besedilo <Pogovor s predsednikom.>: p 71 00 O 62 77 g 53 79 O: 88 79 v 40 87 O 87 94 r 37 77 # s 139 00 p 63 00 r 35 76 E 59 77 ts 78 00 e: 91 87 d 49 94 n 39 97 i 48 102 k 57 00 O 74 75 m 54 70 . 700 00

7.4.2 Oblikovanje stavčne intonacije

Kot intonacijski segment smo vzeli frazo. Pri razgradnji besedila zaenkrat še nismo določali mest poudarkov. To je namreč izjemno zahtevna naloga, ki poleg skladijskega in semantičnega znanja zahteva tudi tankočutno poznavanje tematike govora. Menili smo, da bi z morebitnim nepravilno določenim mestom poudarka močno pokvarili naravnost sintetičnega govora. Če pa bi imeli mesto poudarka podano, bi ga upoštevali preprosto tako, da bi povečali izrazitost tonemskega naglasa v poudarjenem taktu [Srebot 88, str. 151–154].

Ker v frazi ni nič poudarjenega, naglašene zloge v ogrodju stavčne intonacije

višinsko uredimo tako, da je prvi najvišji in vsak naslednji nižji. Zaenkrat uporabljamo linearno padajoče ogrodje, podobno kot Allen [Allen 87, str. 100–107], ki ga nameravamo v prihodnje nadomestiti s kakšnim bolj primernim, morda eksponentnim, kot ga je predlagal Fujisaki [Fujisaki 93], in ki se je v svetu precej uveljavil.

Nato na ogrodje stavčne intonacije superponiramo besedni naglas, ki smo ga predhodno določili. V jezikih, ki ne poznajo tonemskega naglasa, je ta superpozicija preprosta, saj jo dosežejo s prištevanjem t. i. lokalne komponente, ki modelira vpliv poudarka, t. i. globalni komponenti, ki modelira ogrodje stavčne intonacije [Fujisaki 93].

Menimo, da za slovenščino ta princip ni primeren. Za slovenščino je namreč značilna tesna povezanost tonemskega naglasa in stavčne intonacije [Srebot 96], medtem ko besede v angleščini in mnogih drugih jezikih nimajo tonske krivulje, temveč oblikujejo potek krivulje osnovnega tona izključno pod vplivom stavčne intonacije. T. Srebot Rejec nadalje predpostavlja, da je potek stavčne intonacije v angleščini mnogo bolj razgiban kot v slovenščini, medtem ko v slovenščini k tonski razgibanosti bolj prispeva besedni naglas kot stavčna intonacija.

Do podobnega zaključka smo prišli tudi ob preizkusu sintetizatorja govora, ki ga opisujemo v predzadnjem poglavju. Kar 11 od 21 poslušalcev, ki so sodelovali pri preizkusu sintetizatorja, je namreč menilo, da je kakovost sintetiziranega govora brez nastavljene stavčne intonacije enakovredna kakovosti sintetiziranega govora z nastavljeno stavčno intonacijo.

Slovensko stavčno intonacijo slovenska tonematičnost močno zaplete, hkrati pa jo dela slovensko individualno [Toporišič 91, str. 451]. Zato je še posebej težavno določiti njun medsebojni vpliv. Superpozicijo besednega naglasa in stavčne intonacije dosežemo tako, da vrednosti osnovne frekvence, ki jih dobimo po oblikovanju besednega naglasa, skaliramo s faktorjem, ki podaja vpliv stavčne intonacije na osnovno frekvenco v besedi. S tem hkrati povečamo tonski obseg sintetičnega govora.

V nadaljevanju opisujemo, kateri tip stavčne intonacije smo priredili posameznim frazam in kako smo oblikovali intonacijske krivulje, ki jih imenujemo *intonacijski vzorci*.

Vrsto intonacijskega vzorca, ki smo ga priredili posamezni frazi, smo določali izključno na podlagi ločila na koncu fraze [Gros 97e, str. 151].

Težišče intonacije določimo s pomočjo zbirke vprašalnic z označenimi mesti težišča [Gros 97e, str. 151]. Če se v stavku nahaja vprašalnica iz zbirke, je težišče na tej vprašalnici, sicer pa na naglašenem zlogu zadnje besede v frazi.

Grafične ponazoritve značilnih intonacijskih vzorcev, ki smo jih uporabili, so podane v navedeni literaturi [Gros 97e, str. 152, 153].

V tem poglavju opisujemo različne postopke za sintezo govornega signala. Nadaljujemo z opisom postopka pridobivanja zbirke osnovnih enot in združevanja osnovnih enot v govorni signal ob upoštevanju želenih prozodičnih lastnosti za posamezne glasove.

Iz literature poznamo več načinov delitve postopkov za sintezo govora, kot npr. na postopke, ki temeljijo na pravilih, in na postopke, ki temeljijo na združevanju osnovnih enot. Po eni izmed delitev se postopki za sintezo govornega signala delijo v naslednje tri glavne skupine [d'Alessandro 95]: artikulatorni sintetizatorji govora, formantni sintetizatorji govora in sintetizatorji govora, ki temeljijo na združevanju osnovnih enot.

Artikulatorni sintetizatorji govora so fizični modeli, ki so zasnovani na podlagi natančnega opisa fiziologije in akustike govornega trakta [Parthasarathy 92]. So najstarejši sintetizatorji govora.

Značilni parametri takega sintetizatorja govora so položaj in opis kinematike govoril pri govoru. Za posamezne dele govornega trakta se modelirajo prenosne funkcije, kot npr. prenosna funkcija mehanskega modela za prenos vibracij v grlu [Miki 93]. Govorni signal se izračuna na podlagi matematične simulacije pretoka zraka skozi govorni trakt. Pri tem se uravnavajo različni parametri, kot so podgoltni tlak, napetost glasilk in relativni položaj različnih drugih organov [Keller 94].

Ta vrsta sintetizatorjev še ni tržno uspešna. Kljub temu so raziskave v tej smeri zanimive, ker povedo mnogo o tvorjenju glasov pri človeku. Artikulatorni sintetizatorji govora so obširno opisani v navedeni literaturi [Flanagan 72, str. 204–283; Furui 89, str. 25–43; Gabioud 94].

Pri formantni sintezi govora predpostavljamo, da je prenosna funkcija govornega trakta zadovoljivo opisana s formantnimi frekvencami in formantnimi amplitudami [Styger 94]. Sinteza govornega signala temelji na umetni rekonstrukciji formantnih značilnosti. To dosežejo z vzbujanjem filtra (niza resonatorjev) z virom, ki je periodično vzbujanje za zveneče glasove, za nezvенеče glasove pa šum, ki simulira stiskanje gornjega dela govornega trakta. Z dodatkom niza antiresonatorjev lahko simulirajo tudi efekte nosnega trakta, pripornike in zapornike. Podroben opis formantnih sintetizatorjev govora je podan pri Holmesu in Allenu [Holmes 64; Allen 87].

Prednost formantnih sintetizatorjev je v tem, da so njihovi parametri močno povezani z nastajanjem in širjenjem zvoka v govornem traktu. Vendar samodejni postopki za določitev formantnih parametrov še vedno niso dovolj zanesljivi. Modeliranje prehajanja med formanti za različne glasove tudi ni popolno. Fino

uglaševanje parametrov formantnega sintetizatorja je dolgotrajno in mukotržno, poleg tega je treba postopek ponoviti za vsakega novega govorca. Določitev formantov je močno odvisna od jezika, ki ga sintetiziramo, in terja poglobljeno fonetično znanje o danem jeziku.

Kljub navedenim omejitvam so formantni sintetizatorji govora tudi tržno zanimivi. Različice, ki so na voljo, so sposobne sintetizirati govor v mnogih jezikih, in večinoma predstavljajo različne izvedbe znamenitega Klattovega sintetizatorja govora [Klatt 80]. Med predstavnike najbolj znanih formantnih sintetizatorjev se uvrščajo: JSRU [Holmes 64] in MITalk [Allen 87] za angleščino, sistem INRS [Shaughnessy 88] za francoščino ter večjezikovna sistema KTH [Carlson 76] in Infovox [Carlson 82], ki je že leta 1982 obsegal sintezo osmih evropskih jezikov.

Ti postopki sinteze govora temeljijo na obdelavi govornega signala osnovnih enot, ki so vzete iz vnaprej posnetega naravnega govora [d'Alessandro 95]. Zbirka osnovnih enot mora biti sestavljena tako, da odraža fonološke značilnosti danega jezika. Tako je na primer opis glasov zajet v difonih, ki predstavljajo prehode med glasovi, ali celo v večjih govornih enotah, kot so trifoni, polzlogi, zlogi ali celo besede [Shaughnessy 88].

Sintetizator govora ob združevanju govornih enot uporablja postopke obdelave signalov, da zgladi spektralne nezveznosti, ki nastanejo na zlepkih, in hkrati nastavi zahtevane prozodične parametre. Postopki, ki se pri tem uporabljajo, so različni: večpulzno linearno napovedovanje LPC [Atal 82], obdelovanje signalov, sinhrono z osnovno periodo: PSOLA [Moulines 90] in WSOLA [Roelands 93], in kombinacija harmonskih in šumnih modelov: postopek HNS [Laroche 93] in postopek MBR-PSOLA [Dutoit 93a; Dutoit 93b].

Dandanes je na voljo mnogo tržno uspešnih sintetizatorjev govora, ki temeljijo na združevanju osnovnih enot, mnogi med njimi so večjezikovni: sistema CSELT [Balestri 93] in CNET Vox PSOLA [Sorin 94], sistemi HADIFAX [Portele 94], PHRITTS [Meyer 93] in SVOX [Pfister 94] za nemški jezik, *ATR n-talk* [Sagisaka 92], sistem MULTIVOX [Olaszy 91] za vzhodnoevropske jezike in mnogi drugi.

Prvi poskus sinteze slovenskega govora, izveden na Institutu Jožef Stefan, je uporabljal formantna sintetizatorja govora COVOX in LSI, ki sta bila razvita za sintezo angleškega govora. Kakovost sintetiziranega govora je bila temu primerna [Weilguny 93]. Delo na sintezi govora so nadaljevali v smeri difonske sinteze govora [Šef 98].

Naš sistem za sintezo slovenskega govora temelji na združevanju osnovnih govornih enot, difonov, s postopkom TD-PSOLA [Moulines 90]. V nadaljevanju poglavja opisujemo postopek pridobivanja zbirke osnovnih enot in njihovo združevanje.

8.1 Zbirka osnovnih enot

Izbira osnovnih enot ima pri sistemih, ki temeljijo na združevanju osnovnih enot, je zelo pomembna za končno naravnost in razumljivost sintetiziranega govora. Izbiramo lahko med fonemi, alofoni, difoni, polifoni, izgovarjanimi besedami ali izgovarjanimi povedmi [Shaughnessy 88].

Pomembno je, da se akustične razlike med shranjenimi in zahtevanimi izseki in tudi zvočne prekinitve na meji med sosednjimi odseki zmanjšajo na minimum. Difoni kot osnovne enote so najbolj splošno sprejet kompromis med velikostjo osnovne enote in posledično velikostjo zahtevanega spomina v računalniku, med zapletenostjo pravil ter končno kakovostjo sintetiziranega govora. *Difon* je definiran kot del sklopa dveh glasov, in sicer kot sklop med drugim delom prvega glasu in prvim delom drugega glasu. Tako vsebuje prehod med obema glasovoma, ki ga je še posebej težko pravilno napovedati.

Še vedno so mnenja o tem, ali je pri snemanju difonske baze bolje uporabiti smiselne ali nesmiselne besede, ali naj bodo difoni v njih naglašeni ali ne in podobno, deljena [Drullman 93]. Držali smo se priporočil avtorjev algoritma za združevanje osnovnih enot TD-PSOLA [Hamon 88; Moulines 90] in smo difone umestili v sredino brezpomenskih besed, logatomov.

Postopek izgradnje difonske baze smo razdelili v štiri dele: generiranje brezpomenskih besed, logatomov, snemanje izoliranih logatomov, segmentacija ali razčlenjevanje difonov iz izoliranih logatomov in označevanje osnovne periode.

Z umestitvijo difonov v brezpomenske logatome se govorec izogne nenameremu naglaševanju pri snemanju besed. Logatome sestavimo tako, da se difon nahaja v čim bolj nevtralnem fonetičnem okolju.

Logatome smo sestavili iz treh delov: v sredini logatoma se nahajata glasova, iz katerih izrežemo difon, levo in desno od njiju sta zloga, ki poskrbita, da se na osrednjih glasovih ne odražata začetni in končni prehodni pojav. Prehodni pojav se namreč pogosto izraža s spremembo glasnosti in osnovne frekvence, te spremembe pa so v difonu nezaželene.

Sprva moramo izbrati zbirko glasov, ki jih želimo sintetizirati. Pri določanju števila difonov se omejimo na sklope glasov, ki so v danem jeziku dopustni [Srebot 75], [Toporišič 91]. Ker pa mora biti sintetizator govora robusten in mora zadnji modul v sistemu vedno najti difon, zahtevan v zaporedju glasov fonetične transkripcije, smo dodali tudi nekaj glasovnih sklopov, ki po pravilih v slovenskem jeziku sicer niso dopustni. Te sklope sintetizator govora uporablja zelo redko, najbolj pogosto pa se pojavijo na stikih med besedami, kjer je bila fonetična transkripcija vhodnega besedila nepravilno določena.

Pri tvorjenju logatomov se držimo predvsem osnovnega zlogotvornega načela, ki podobno kot za druge jezike velja tudi za slovenski jezik, tj. zgradbe NZVZN nezvočnik-zvočnik-samoglasnik-zvočnik-nezvočnik [Srebot 75a; Srebot 75b; Srebot 92]. To pravilo uporabimo za umeščanje samoglasniških delov difonov, so-

glasnike pa uvrstimo med dva samoglasnika VCV, razen v primeru, ko difon predstavlja soglasniški sklop.

Pravila, ki smo jih uporabili pri sestavljanju logatomov, so izčrpno opisana pri Grencu [Grenc 95, str. 28–31]. Predvsem smo morali biti pazljivi v primeru alofonskih glasovnih variant določenega fonema, da smo jih uvrstili v pravilno glasovno okolje, kot npr. pri mehčanem [l'] in [n'], pri mehkonebnem [N], pri zobnoustničnem [F] in pri izgovornih različicah črke <v>. Omenjamo črko <v> zato, ker fonemu /v/ težko pripišemo vse izgovorne različice, ki nastanejo ob izgovarjavi te črke [Srebot 81].

Končna zbirka je obsegala 1028 različnih logatomov. V njej smo zajeli glasovne sklope, ki ob izbranem glasovnem izboru nastopijo v slovenščini.

Izbrano zbirko logatomov smo posneli in govor povzorčili ob frekvenci vzorčenja 16 kHz. Logatome je ponovno izgovarjal govorec F. M. Govorca smo prosili, da logatome izgovarja neprizadeto, nenaglašeno in čim bolj monotono. Pred izgovarjavo vsakega logatoma je zaslišal pisk z osnovno frekvenco, ki je približno ustrezala sredini njegovega glasovnega obsega.

Ker se je govorec močno trudil pravilno izgovarjati nesmiselne besede, jih je nehote izgovarjal prepočasi, tako da dobljeni difoni trajajo predolgo. Zato predlagamo, da se ob naslednjem snemanju difonske baze že sproti budno spremljata trajanje in osnovna frekvenca glasov v difonu.

Posnete logatome smo ročno označili. Pridobivanje difonskega signala iz logatoma smo razdelili v tri korake: groba segmentacija logatoma, fina segmentacija difona in označevanje sredine prehoda med glasovoma znotraj difona.

V prvem koraku smo iz logatoma izrezali grobo okolico glasovnega para, v katerem se je nahajal želeni difon. Sledilo je določanje mej difona. Mejo difona smo postavili približno na sredino časovnega poteka za posamezni glas v difonu. Le pri zapornikih smo pazili na to, da je je zapornik kot drugi glas v difonu obsegal le zaporo, kot prvi glas v difonu pa del zapore in celotno odporo. Nazadnje smo označili še mejo med obema glasovoma v difonu.

Pri razčlenjevanju govornega signala logatomov na difone smo si pomagali s programskim paketom, ki je bil razvit v Laboratoriju za umetno zaznavanje na Fakulteti za elektrotehniko v Ljubljani [Dobrišek 96b]. Programski paket omogoča predvajanje izbranih delov govornega signala, njihovo predstavitev v časovnem in v frekvenčnem prostoru ter postavljanje oznak na izbrane dele signala.

Na prehodih med difoni morata biti tako amplituda kot tudi spekter zvezna in razmeroma stalna [Shaughnessy 88]. Nenadne amplitudne spremembe so dopustne le na prehodih pri zapornikih, zlitnikih in nosnikih, ker takrat pride do zapore oziroma odpore govornega trakta. Da bi preprečili nezveznosti v amplitudi ob prehodih med difoni, smo amplitudo na prehodih med difoni normirali. Na amplitudo v sredini glasu, kar ustreza meji difona, fonetično okolje glasu razmeroma malo vpliva, zato smo amplitudo na mejah difonov prilagodili povprečni vrednosti amplitude na meji za določen glas.

Postopek, ki smo ga uporabili za končno sintezo govornega signala, zahteva označitev osnovnih period v difonskem signalu. Pri postavljanju oznak osnovne periode smo morali biti dosledni. Za karakteristično točko smo izbrali ničlo pred globalnim minimumom v periodi in ta točka je predstavljala začetek periode. Z opazovanjem poteka signala smo določili preostale osnovne periode tega glasu ter osnovne periode drugega glasu v difonu, če je bil le-ta zveneč. Pri postavljanju oznak za osnovno periodo smo opazovali predvsem frekvenčni potek govornega signala difona.

Označevanje osnovnih period samoglasnikov je bilo preprosto. Manjše težave so se pojavile pri označevanju zvenečih soglasnikov [j], [l], [m], [n], [v], [z] in [Z] zaradi na oko neizrazite osnovne periode v signalu. Šum, ki je superponiran glasu z nizko amplitudo, pri glasu [Z] še bolj oteži določanje osnovne periode. Poseben primer je bil tudi glas [r].

Ročno označevanje difonske baze je dolgotrajen in utrujajoč postopek. Zato se v zadnjem času vse bolj uveljavljajo postopki za samodejno pridobivanje zbirk difonov [Cosi 91]. Ti postopki večinoma temeljijo na prikritih Markovovih modelih [Taylor 91; Ljolje 93; Donovan 95] ali pa na dinamičnem prileganju govornega signala z referenčnim [Ottesen 93]. Predvsem je zanimiv pristop dinamičnega časovnega prileganja s sintetiziranim besedilom, kjer so meje med glasovi vnaprej poznane [Portele 96; Campbell 96], in ki je bil nedavno uporabljen tudi za samodejno razčlenjevanje slovenskega govora [Dobrišek 97].

Za samodejno segmentacijo logatomov smo uporabili samodejni postopek [Gros 96c], ki temelji na prikritih Markovovih modelih [Ipšič 96]. Vendar se je izkazalo, da tako določanje mej med glasovi ni dovolj zanesljivo in je tako dobljene meje treba ročno popraviti. Za samodejno določanje oznak osnovne periode smo uporabili algoritem SRPD [Medan 91], ki smo ga posebej prilagodili za postavljanje oznak na karakteristična mesta v periodi [Senčar 96; Sovdat 96].

8.2 Združevanje osnovnih enot

Za združevanje osnovnih enot smo izbrali algoritem TD-PSOLA, ker je preprost, hiter in zagotavlja visoko kakovost sintetiziranega govora. Algoritem PSOLA (angl. pitch synchronous overlap add) [Hamon 87; Hamon 88; Moulines 90] je neparametrična metoda in za razliko od LPC metode [Atal82] ne uporablja nobenega modela govornega signala, razen dejstva, da je govorni signal na zvenečih delih periodičen. Na ta način se kakovost shranjenih delov govora v celoti ohrani.

Sinteza PSOLA temelji na razčlembi signala v zaporedje prekrivajočih se kratkotrajnih signalov, dobljenih z oknenjem izvornega signala [Hamon 87]. Omožja kakovostne frekvenčne in časovne spremembe neposredno na izvornem signalu, vsaj za zmerne spremembe prozodičnih parametrov.

Sprememba govornega signala je možna tako v frekvenčnem prostoru z uporabo diskretne Fourierove transformacije, frequency-domain PSOLA, imenovana FD-PSOLA, kot tudi neposredno v časovnem prostoru, time-domain PSOLA, imenovana TD-PSOLA. FD-PSOLA omogoča veliko fleksibilnost in spremenljivost spektralnih karakteristik govornega signala, medtem ko je TD-PSOLA zanimiva za izvedbo sistemov sinteze govora v stvarnem času. Prav zaradi računске nezahtevnosti algoritma je slednji postopek še posebej zanimiv.

Če spremembe prozodičnih parametrov presežejo določen okvir, je ob prevelikem podaljšanju nezvenečih glasov moč zaznati slišne ostanke harmoničnih komponent, kar je posledica rednega ponavljanja nezvenečih kratkočasovnih signalov. To pomanjkljivost naj bi odpravil mešani model. Algoritem LP-PSOLA uporablja za spremembo prozodičnih parametrov metodo TD-PSOLA. Z inverznim filtrom dodaja še spektralno informacijo. Uporaba te tehnike omogoča učinkovito spektralno glajenje in spremembo ovojnice na mestih združevanja.

Zaradi računске nezahtevnosti in preprostosti algoritma smo se odločili, da v prvi različici sintetizatorja govora uporabimo postopek združevanja enot z algoritmom TD-PSOLA [Gros 95], [Šef 95a; Gros 96a]. Algoritem omogoča sočasen nadzor nad osnovno frekvenco in nad trajanjem sintetiziranega signala. Ker deluje sinhrono z osnovno periodo govornega signala, je bilo treba v zbirki difonov predhodno označiti osnovne periode.

Postopek sinteze govora TD-PSOLA sestavljajo trije koraki [Moulines 90]:

1. razgradnja izvornega govornega signala, sinhrono z osnovno frekvenco: sprva izvorni govorni signal difonov razdelimo na množico prekrivajočih se kratkočasovnih signalov, ki jih dobimo z oknenjem izvornega signala; tako dobimo vmesno neparametrično predstavitev signala;
2. spremembe na vmesni predstavitvi signala, sinhrono z osnovno frekvenco in
3. sinteza novega signala iz vmesnih predstavitev signala, sinhrono z osnovno frekvenco.

Posamezni koraki postopka so natanko opisani v navedeni literaturi [Šef 95; Gros 97e, str. 160–169].

S preizkusi kakovosti sintetizatorjev govora se ocenjujeta predvsem naravnost in razumljivost sintetiziranega govora. Opis preizkusov najbolj znanih sintetizatorjev govora je zbran pri Polsu [Pols 89].

Razumljivost sintetiziranega govora so merili na več nivojih, na segmentnem nivoju (modificirani test rimanja, angl. *modified rhyme test*), na nivoju besed in na stavčnem nivoju (test s semantično nenapovedljivimi stavki SUS [Benoit 90], angl. *semantically unpredictable sentences*) [Jekosch 93; Weilguny 93, Gros 95]. Tovrstni preizkusi sintetizatorja govora so težko ponovljivi, saj pogoji, v katerih so bili izvedeni, navadno niso bili natančno opisani [Pols 94].

Preizkusi s t. i. diagnostičnimi podmnožicami so namenjeni primerjavi med različnimi pristopi pri nastavljanju določenega parametra sintetizatorja govora in izbiri najbolj primerne [Bezooijen 90]. Z njimi lahko tudi ugotavljamo šibke točke sintetizatorja govora, npr. iskanje glasov, ki so nerazločno izgovarjani.

V zadnjem času se vse bolj uveljavljajo postopki za subjektivno ocenjevanje kakovosti sintetiziranega govora, ki so standardizirani pri mednarodnem telekomunikacijskem združenju ITU [Niebank 93; Niebank 95]. Izhajajo iz standardov, ki služijo za ocenjevanje kakovosti govora pri telefonskem prenosu, še zlasti iz priporočila P.80 (ITU Recommendation P.80 – Subjective opinion tests) [ITU 88]. Tak način ocenjevanja je nadvse primeren pri ugotavljanju, ali je določen sintetizator govora uporaben v točno določen namen [Pols 95].

Kakovost sintetiziranega govora se ocenjuje z opisnimi spremenljivkami, ovrednotenimi po ITU lestvici MOS (angl. *mean opinion score*) [ITU 93]. Lestvica MOS se uporablja pri subjektivnem vrednotenju kakovosti govora, predvsem v telefoniji. Lestvica MOS je večstopenjska. Najbolj pogosto ima pet stopenj, ki si sledijo od najslabše, ocena 1, do najboljše, ocena 5.

Da bi ugotovili kakovost, naravnost in možnost uporabe umetno sintetiziranega govora, smo v Laboratoriju za umetno zaznavanje organizirali preizkus sintetizatorja govora, pri katerem je sodelovalo 21 poslušalcev v starostnem razponu med 19 in 45 leti. 10 poslušalcev je bilo ženskega spola in 11 moškega spola, tretjina poslušalcev je imela srednjo izobrazbo, dve tretjini visoko, dva poslušalca nista bila slovenskega rodu, 18 poslušalcev pred preizkusom še ni slišalo sintetizatorja govora. Preskus je bolj natančno opisan v diplomskem delu [Mazej 97].

Del preizkusa smo zasnovali po priporočilih mednarodne zveze za telekomunikacije ITU-TSS [ITU 93; ITU 94], v preostalem delu preizkusa smo želeli dobiti splošno oceno sintetizatorja govora in oceno posameznih nastavitvev sintetizatorja govora z diagnostičnimi podmnožicami.

Celoten preizkus je po mednarodnih priporočilih trajal manj kot eno uro. Preizkus smo ponovili štirikrat. Vsakokrat je pri njem sodelovalo 5 ali 6 poslušalcev. Poslušalci so ocene vnašali na posebne pole. Sprva smo poslušalcem predvajali uvodno sporočilo, tako da so se privadili sintetiziranemu govoru.

Preizkus kakovosti sintetizatorja govora za slovenski jezik smo razdelili na tri dele [Mazej 97]: ocenjevanje primernosti sintetizatorja govora za dano področje uporabe, primerjavo kakovosti sintetičnega govora s pošumljenim naravnim govorom in opisno ocenjevanje kakovosti sintetiziranega govora.

Prvi del preizkusa smo zasnovali po napotkih iz mednarodnega priporočila ITU-P.85 (ITU Recommendation P.85 – A method for subjective performance assessment of the quality of speech voice output devices) [ITU 94], ki govori o ocenjevanju kakovosti sintetičnega govora v interaktivnih govornih odzivnih sistemih. Poslušalci v vnaprej pripravljene obrazce vpisujejo določene informacije, ki so jih slišali in ki se nanašajo na konkretno področje uporabe. Pri izbiri področja uporabe smo se odločili za uporabo sintetizatorja v samodejnem telefonskem odzivniku za podajanje informacij o letalskih povezavah.

Prvi del preizkusa je pokazal, da je sintetični govor poslušalcem razumljiv. Poslušalce smo povprašali, ali se jim zdi, da bi bil sintetizator govora v tej obliki primeren za samodejno podajanje informacij po telefonu. Več kot dve tretjini poslušalcev sta menili, da bi bil primeren, manj kot ena tretjina je menila, da ne bi bil primeren. Kar nekaj poslušalcev, ki so obkrožili 'ne', je po preizkusu neuradno pripomnilo, da so tako oceno dali predvsem zato, ker se jim zdi neprimereno, da bi stroji ljudem odvzemali delovna mesta, in ne zato, ker bi se jim zdela kakovost sintetiziranega govora prenizka.

Za ocenjevanje kakovosti govornih naprav ITU predlaga postopek, kjer poslušalci primerjajo sintetični govor z referenčno enoto, ki naravni govor popači s šumom. Referenčna enota, imenovana MNRU, je predstavljena v Modri knjigi združenja ITU kot priporočilo P.81 (ITU-T Recommendation P.81 Telephone transmission quality subjective opinion tests – Modulated Noise Reference Unit – MNRU) [ITU 93]. Govorni signal je časovno pomnožen z Gassovim šumom. Šum, moduliran z govorom, se prišteje izvirnemu govornemu signalu tako, da je doseženo predpisano razmerje signal šum. Dobljeni signal smo poslali skozi ozkopasovni filter, ki prepušča frekvence med 100 in 3400 Hz tako kot telefonska linija. Uporabili smo Chebishev filter osmega reda, ki smo ga načrtali s pomočjo študentske različice programskega paketa za digitalno obdelavo signalov MONARCH.

V drugem delu testa so poslušalci kakovost sintetičnega govora uvrstili med naravni govor, pošumljen z moduliranim Gaussovimi šumom v razmerju signal : šum 5 dB in 10 dB.

V tretjem delu testa so poslušalci ocenjevali različne nastavitve parametrov sintetizatorja govora. Nanaša se na priporočilo ITU-P.85 (ITU Recommendation P.85-Telephone transmission quality subjective tests – a method for subjective

performance assessment of the quality of speech voice output devices) [ITU 94].

V tem delu preizkusa smo sprva ocenili splošno kakovost sintetiziranega govora. Preizkus smo nadaljevali z ocenjevanjem diagnostičnih podmnožic besedil, ki so bila tvorjena z različnimi nastavitvami sintetizatorja govora, kot to predlaga Bezooijen [Bezooijen 90].

Natančni opisi vseh treh delov preizkusa ter razlaga rezultatov so podani v navedeni literaturi [Gros 97e, str. 170–180].

Delo obravnava pregled stanja jezikoslovja in tehnike na področju samodejnega tvorjenja govora iz besedil. Nadalje podaja napotke, kako izgraditi sintetizator govora za slovenski jezik ter navaja številčne vrednosti parametrov, s katerimi krmilimo sintetizator govora.

Opisali smo meritve prozodičnih parametrov, ki jih je bilo treba opraviti za slovenski govor. Izgradnja sintetizatorja govora za slovenski jezik je bila zasnovana modularno. Sam postopek sinteze govornega signala temelji na združevanju vnaprej posnetih osnovnih govornih enot ob upoštevanju zahtevanih sprememb v trajanju in v osnovni frekvenci za posamezne glasove. Sintetizator govora na izhodu predvaja tekoč govor, ki je prozodično razgiban.

Prvi korak postopka sinteze govora je predobdelava besedila. Sledi določanje fonetičnega prepisa besed. S statistično analizo obsežne zbirke slovenskih besedil različnih zvrsti smo prišli do seznama 40.000 najbolj pogostih slovenskih besed. Te besede pokrijejo preko 80 % besed, ki nastopijo v povprečnem slovenskem besedilu. Izgovarjave besed smo za kanonične oblike besed določili ročno. Ko se izgovarjava besede iz slovarja uporabi pri grafemsko-fonetični pretvorbi besedila, se nad njo naknadno izvede preslikava s pravili koartikulacije med besedami.

Slovar izgovarjav je zaradi lažjega dodajanja novih slovarskih gesel razdeljen v sedem podslovarjev: slovar kratic, slovar števil, slovar lastnih imen, slovar besed tujega izvora, slovar pogostih besednih zvez ali kolokacij, slovar okrajšav in slovar preostalih pogostih besed. Ker je za slovenski jezik značilen premični naglas, je naglasno mesto brez slovarja izgovarjav za vse pregibne oblike izvorne besede težko pravilno napovedati. Ker so najbolj pogoste oblike besed zajete že v slovarju izgovarjav, za preostale oblike določamo naglasno mesto samodejno. Pri tem upoštevamo pravila, do katerih smo prišli ob statistični analizi razporejanja naglasnega mesta po zlogih v raznozložnih besedah v slovarju pogostih izgovarjav.

Če besede ne najdemo v slovarju izgovarjav, ji je treba fonetični prepis določiti samodejno. Sestavili smo zbirko pravil s preko 140 kontekstno odvisnimi pravili, ki predpisujejo preslikavo grafemskega zapisa slovenskega besedila v fonetični prepis. V pravilih je zajet tudi vpliv koartikulacije na mejah med sosednjimi besedami. Pravila smo preskušali in dopolnjevali na obsežni zbirki pogostih slovenskih besed z njihovimi izgovarjavami.

Sledilo je nastavljanje prozodičnih parametrov govora. Pred tem smo morali izvesti nekaj meritev prozodičnih parametrov v naravnem govoru. Na obsežni bazi izgovarjav tekočega govora, kjer je govorec isto besedilo izgovarjal s tremi raz-

ličnimi hitrostmi govora, smo merili povprečna trajanja glasov ter trajanje raznozložnih taktov v različnih legah v frazi. Tovrstne raziskave za slovenski jezik še niso bile opravljene. Za višjenivojsko govorno enoto smo namesto besede izbrali takt, kot skupek besed in od njih naglasno odvisnih breznaglasnic, ker bi sicer kratke enozložne breznaglasnice pokvarile splošnost rezultatov pri določanju tempa artikulacije.

Zakovitosti v spremembah trajanja za posamezne glasove smo uspeli posplošiti na večje glasovne skupine. To pomeni, da se ob spremenjeni hitrosti govora trajanje glasov, ki pripadajo isti glasovni skupini, spreminja na podoben način. Domnevamo, da do tega pride tudi zaradi upiranja govoril spremembi hitrosti govora. Odvisnost tempa artikulacije od lege takta v frazi smo merili za takte z različnim številom zlogov. Odvisnost tempa artikulacije od števila zlogov v taktu smo uspeli opisati z matematičnim modelom.

Trajanje glasov smo nastavljali z dvostopenjskim modelom. V prvem koraku za vsak glas določimo njegovo intrinzično trajanje na podlagi inherentnega trajanja glasu, položaja glasu v besedi, strukture zloga, v katerem se glas nahaja, in njegovega glasovnega okolja. V drugem koraku določimo ekstrinzično trajanja takta glede na želeno hitrost govora, število zlogov v taktu in položaj takta v frazi. Za usklajevanje intrinzičnega trajanja glasov in ekstrinzičnega trajanja taktov smo razvili tri postopke, ki upoštevajo različno spreminjanje glasov ob spremembi hitrosti govora, metodo translacije, metodo skaliranja in proporcionalno metodo. Uspešnost razvitih postopkov pri napovedovanju trajanja glasov smo ocenili s primerjavo napovedanih trajanj glasov s trajanji glasov, vzetimi iz naravnega govora pri različnih hitrostih govorjenja.

Oblikovanje tonemskega naglasa taktov je sestavljeno iz dveh delov: iz določanja tipa tonemskega naglasa in iz izbire primernega značilnega naglasnega vzorca za izbrani tip tonemskega naglasa. Pri določanju tipa tonemskega naglasa smo izvedli mnogo slušnih preizkusov in prišli do zaključkov za mešano izbiro naglasov, podkrepjenih z jezikoslovno teorijo. Omejitve skokov pri značilnih naglasnih vzorcih za izbrani tip tonemskega smo fino nastavili s slušnimi preizkusi.

Za izbrani izbor slovenskih glasov smo definirali pravila, po katerih smo sestavili logatome, tako da so ti vsebovali izbrani difon v fonetično čim bolj nevtralni legi. Končno zbirko logatomov smo omejili z določanjem sklopov dveh glasov, ki so v slovenskem fonološkem sistemu dopustni. Ker je ročna segmentacija zbirke difonov dolgotrajen in težaven postopek, smo ga skušali vsaj delno avtomatizirati.

Sintetizator slovenskega govora smo vgradili v samodejni sistem za podajanje informacij o letalskih poletih [Gros 97]. Razumljivost in naravnost sintetiziranega govora smo ocenili z obsežnim preizkusom, ki je bil pripravljen po mednarodnih priporočilih za preverjanje kakovosti sintetiziranega govora.

Načrte za prihodnje delo lahko razdelimo v več delov. Pri določanju fonetičnega

prepisa vhodnega besedila priporočamo, da se slovar pogostih izgovarjav še dodatno razširi z novimi besedami. Samodejna skladenjska in pomenska razgradnja vhodnega besedila omogočata napovedovanje poudarka v frazi.

Pri določanju prozodičnih parametrov se je treba posvetiti predvsem nastavljanju osnovne frekvence. Pri oblikovanju tonemskega naglasa bi bilo zanimivo upoštevati še več vzorcev osnovne frekvence. Vrsto tonemskega naglasa za najbolj pogoste besede bi lahko shranili v slovar izgovarjav. Tudi pri oblikovanju poteka stavčne intonacije lahko upoštevamo več intonacijskih vzorcev.

Nazadnje predlagamo razširitev obstoječe difonske baze z novimi difoni, še zlasti z difoni, ki vsebujejo dele dvoglasnikov, in pa z bolj sredinsko izgovorjenimi samoglasniki. Izgovarjava nenaglašanih samoglasnikov je namreč preveč izrazita, na kar nas je prijazno opozorila T. Srebot Rejec.

- [Abadjieva 93] E. Abadjieva, I. R. Murray, J. L. Arnott. *Applying analysis of human emotional speech to enhance synthetic speech*. Proceedings of the EUROSPEECH'93. Berlin. str. 909–1002. 1993.
- [Abe 92] M. Abe, H. Sato. *Two-stage control model using syllable based F0 units*. Proceedings of the ICASSP'92. San Francisco. 1992.
- [Adamson 96] M. J. Adamson, R. I. Dampier. *A recurrent network that learns to pronounce English text*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Ainsworth 89] W. A. Ainsworth, B. Pell. *Connectionist architectures to a text-to-speech system*. Proceedings of the EUROSPEECH'89. str. 125–128. 1989.
- [Allen 87] J. Allen, M. S. Hunnicutt, D. H. Klatt. *From text to speech: The MITalk system*. Cambridge University Press. Cambridge. 1987.
- [d'Alessandro 95] C. d'Alessandro, J. S. Lienard. *Synthetic speech generation. Survey of the State of the Art in Human Language Technology*. str. 196–201. 1995.
- [Andersen 96] O. Andersen, R. Kuhn, A. Lazarides. P. Dalsgaard, J. Haas, E. Nöth. *Comparison of two tree-structured approaches for grapheme-to-phoneme conversion*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Atal 82] B. S. Atal, J. R. Remde. *A new model of LPC excitation for producing natural sounding speech at low bit rates*. Proceedings of the ICASSP'82. Pariz. str. 614–617. 1982.
- [Bachenko 90] J. Bachenko, J. Fitzpatrick. *A computational grammar of discourse –neutral prosodic phrasing in English*. Computational Linguistics. (16). str. 155–170. 1990.
- [Bagshaw 93] P. C. Bagshaw, S. M. Hiller, M.A. Jack. *Enhanced pitch tracking and the processing of F0 contours for computer aided intonation teaching*. Proceedings of the EUROSPEECH'93. Berlin. str. 1003–1006. 1993.
- [Bagshaw 94] P. C. Bagshaw. *Automatic Prosodic Analysis for Computer Aided Pronunciation Teaching*. Doktorska disertacija. Univerza v Edinburgu. Velika Britanija. 1994.
- [Bakran 84] J. Bakran. *Model vremenske organizacije hrvatskoga standardnog govora*. Doktorska disertacija. Sveučilište u Zagrebu. 1984.
- [Balestri 93] M. Balestri, S. Lazzaretto, P. L. Salza, S. Sandri. *The CSELT system*

- for Italian text-to-speech synthesis. Proceedings of the EURO-SPEECH'93. Berlin. str. 2091–2094. 1993.
- [Bamakidis 87] S. Bamakidis, G. Carayannis. *Phonemia – a phoneme transcription system for speech synthesis in modern Greek*. Speech Communication. (6). str. 159–169. 1987.
- [Barbosa 93] P. Barbosa, G. Bailly. *Generation and evaluation of rhythmic patterns for text-to-speech synthesis*. Proceedings of the ESCA Workshop on Prosody. Lund. Švedska. str. 66–69. 1993.
- [Bartkova 87] K. Bartkova, C. Sorin. *A model of segmental duration for speech synthesis in French*. Speech Communication. (6). str. 245–260. 1987.
- [Batliner 93] A. Batliner, R. Kompe, A. Kiessling, E. Nöth, H. Niemann, U. Kilian. *The prosodic marking of accents and phrase boundaries: expectations and results*. Proceeding of the NATO ASI Conference. Bubio. str. 89–92. 1993.
- [Belhoula 94] K. Belhoula, V. Kraft, A. Rinscheid, H. W. Ruhl. *Extension of a TTS system to rule-based pronunciation of names*. Proceedings of the CRIM/FORWISS Workshop on Progress and Prospects of Speech Research and Technology. Uredili H. Niemann in drugi. München. str. 249–252. 1994.
- [Benedik 96] M. Benedik. *Proučevanje tempa artikulacije govorca pri različnih hitrostih govora*. Diplomsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1996.
- [Benguerel 86] A. P. Benguerel, J. D'Arcy. *Time-warping and the perception of rhythm in speech*. Journal of Phonetics. (14). str. 231–246. 1996.
- [Benoit 90] C. Benoit. *An intelligibility test using semantically unpredictable sentences: towards the quantification of linguistic complexity*. Speech Communication. (9). str. 293–304. 1990.
- [Bezljaj39] F. Bezljaj. *Oris slovenskega knjižnega izgovora*. Znanstveno društvo. Ljubljana. 1939.
- [Bezoojen 94] R. van Bezooijen, L. C. W. Pols. *Evaluation of text-to-speech conversion in Dutch*. Analysis and Synthesis of Speech. Mouton de Gruyter. Berlin – New York. str. 339–350. 1993.
- [Black 96] A. W. Black, A. H. Junt. *Generating F0 contours from ToBI labels using linear regression*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Bosch 95] L. F. M. ten Bosch. *Automatic classification of pitch movements via MLP-based estimation of class probabilities*. Proceedings of the ICASSP'95. Adelaide. 1995.
- [Breen 95] A. P. Breen. *A simple method of predicting the duration of syllables*. Proceedings of the EUROSPEECH'95. Madrid. str. 595–598. 1995.

- [Breen 96] A. P. Breen, E. Bowers, W. Welsh. *An investigation into the generation of mouth shapes for a talking head*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Bruce 77] G. Bruce. *Swedish word accents in sentence perspective*. Gleerup. Lund. 1077.
- [Burg 96] M. Burg. *Slovar izgovorjav za sintetizator slovenskega govora*. Diplomsko delo. Fakulteta za elektrotehniko. Univerza v Ljubljani. 1996.
- [Campbell 90] W. N. Campbell. *Analog I/O nets for syllable timing*. Speech Communication: Special Issue on Neural Nets and Speech. (9). str. 57–61. 1990.
- [Campbell 91] W. N. Campbell, S.D. Isard. *Segment durations in a syllable frame*. Journal of Phonetics: Special issue on speech synthesis. (19). str. 37–47. 1991.
- [Campbell 93] W. N. Campbell. *Predicting syllable durations for accomodation within a syllable-level timing network*. Proceedings of the EUROSPEECH'93. Berlin. str. 1082–1085. 1993.
- [Campbell 96] W. N. Campbell. *Autolabelling Japanese TOBI*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Carlson 76] R. Carlson, B. Grandstöm. *A text-to-speech system based entirely on rules*. Proceedings of the ICASSP'76. Philadelphia. str. 686–688. 1976.
- [Carlson 82] R. Carlson, B. Granstrom, S. Hunnicut. *A multi-language text-to-speech module*. Proceedings of the ICASSP'82. Pariz. str. 1604–1607. 1982.
- [Chen 92] S. Chen, S. Hwang, C. Tsai. *A first study on neural net based generation of prosodic and spectral information for Mandarin text-to-speech*. Proceedings of the ICASSP'92. San Francisco. 1992.
- [Chou 96] F. Chou, C. Tseng, L. Lee. *Automatic generation of prosodic structure for high quality Mandarin speech synthesis*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Cohen 67] A. Cohen, J. t'Hart. *On the anatomy of intonation*. Lingua. (19). 1967.
- [Cohen 96] A. D. Cohen. *Embedding prior knowledge in a text-to-speech system using data driven methods*. Tehnično poročilo. Univerza v Readingu. 1996.
- [Cole 95] R. A. Cole. *Spoken output technologies. Survey of the State of the Art in Human Language Technology*. str. 189. 1995.
- [Collier 90] R. Collier. *On the perceptual analysis of intonation*. Speech Communication. (9). str. 443–451. 1990.
- [Cosi 91] P. Cosi, D. Falavigna, M. Omologo. *A preliminary statistical eva-*

- luation of manual and automatic segmentation discrepancies.* Proceedings of the EUROSPEECH'91. str. 693–696. 1991.
- [Crystal 69] D. Crystal. *Prosodic systems and intonation in English.* Cambridge Studies in Linguistics. (1). Cambridge University Press. Cambridge. 1969.
- [Crystal 90] T. Crystal, A. House. *Articulation rate and the duration of syllables and stress groups in connected speech.* Journal of the Acoustical Society of America. (88). str. 101–112. 1990.
- [Cummins 96] F. Cummins. *Rhythmic constraints in English stress timing.* Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Dalsgaard 89] P. Dalsgaard. *Semi-automatic phonemic labelling of speech data using a self-organising neural network.* Proceedings of the EUROSPEECH'89. str. 541–544. 1989.
- [Damper 96] R.I. Damper, J.F.G. Eastmond. *Pronouncing text by analogy.* Proceedings of the COLING'96. Copenhagen. str. 268–273. 1996.
- [Dedina 91] M.J. Dedina, H.C. Nusbaum. *Pronounce: a program for pronunciation by analogy.* Computer Speech and Language. (5). str. 55–64. 1991.
- [Dobnikar 95] A. Dobnikar, J. Bakran. *New approach for Slovene text-to-speech synthesis.* Proceedings of MIPRO'95. Opatija. str. 312–318. 1995.
- [Dobnikar 96] A. Dobnikar. *Modeling segment intonation for Slovene TTS system.* Proceedings of the ICSLP'96. (3). Philadelphia. str. 1864–1867. 1996.
- [Dobrišek 96a] S. Dobrišek, J. Gros, F. Mihelič, K. Pepelnjak, I. Ipšič. *GOPOLIS: Slovenian speech database of spoken flight information queries.* Proceedings of the 3rd Slovenian-German and 2nd SDRV Workshop on Speech and Image Understanding. Ljubljana. str. 37–46. 1996.
- [Dobrišek 96b] S. Dobrišek, Z. Kačič, J. Gros, B. Horvat, F. Mihelič. *Pobuda za standardizacijo računalniškega simbolnega fonetičnega zapisa slovenskega govora.* Zbornik ERK'96. (B). str. 247–250. Portorož. 1996.
- [Dobrišek 97] S. Dobrišek, J. Gros, F. Mihelič, N. Pavešič. *Automatic segmentation and labelling for the GOPOLIS speech database.* Proceedings of the 2nd SQEL Workshop on Multi-Lingual Information Retrieval Dialogs. Plzen. str. 80–83. 1997.
- [Donovan 95] R. E. Donovan, P. C. Woodland. *Improvements in an HMM-based speech synthesiser.* Proceedings of the EUROSPEECH'95. Madrid. str. 573–576. 1995.
- [Dreyfus 94] H. L. Dreyfus. *What Computers Still Can't Do.* MIT Press. Cambridge. Massachusetts. 1994.
- [Drullman 93] R. Drullman, R. C. Collier. *Speech synthesis with accented and*

- unaccented diphones*. Analysis and Synthesis of Speech. uredila V.J. van Heuven in L.C.W. Pols. Mouton de Gruyter. str. 147–155. 1993.
- [Dutoit 93a] T. Dutoit. *High Quality Text-to-Speech of the French Language*. Doktorska disertacija. Univerza v Monsu. 1993.
- [Dutoit 93b] T. Dutoit, H. Leich. *MBR-PSOLA: Text-to-speech synthesis based on MBE resynthesis of the segments database*. Speech Communication. (13). str. 432–440. 1993.
- [Epitropakis 93a] G. Epitropakis, D. Tambakas, N. Fakotakis, G. Kokkinakis. *Duration modelling for the Greek language*. Proceedings of the EUROSPEECH'93. Berlin. str. 1995–1998. 1993.
- [Epitropakis 93b] G. Epitropakis, N. Yiourgalis, G. Kokkinakis. *Prosody control of TTS-systems based on linguistic analysis*. Proceedings of the EUROSPEECH'93. Berlin. str. 1999–2002. 1993.
- [Eppinger 86] B. Eppinger, E. Herter. *Sprachverarbeitung*. Hanser Verlag. 1986.
- [Erjavec 92] T. Erjavec, S. Kodrič, I. Božič. *RBSJ – Računalniško Besedišče Slovenskega Jezika*. Delovno poročilo. IJSDP–6355. 1992.
- [Erjavec 97] T. Erjavec. *Unification, inheritance and paradigms in the morphology of natural languages*. Doktorska disertacija. Fakulteta za računalništvo in informatiko. Univerza v Ljubljani. 1997.
- [Erpič 94] T. Erpič. *Pretvorba besedil v foneme*. Diplomsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1994.
- [Eskenazi 93] M. Eskenazi. *Trends in speaking styles research*. Proceedings of the EUROSPEECH'93. Berlin. str. 501–509. 1993.
- [Falavigna 90] D. Falavigna, M. Omologo. *A DTW based approach to the automatic labelling of speech according to the phonemic transcription*. Proceedings of the EUSIPCO'90. str. 1139–1142. 1991.
- [Fant 96] G. Fant, A. Kruckenberg. *On the quantal nature of speech timing*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Ferreira 93] F. Ferreira. *Creation of prosody during sentence production*. Psychological review. (2). str. 233–253. 1993.
- [Flanagan 72] J. L. Flanagan. *Speech Analysis, Synthesis and perception*. Springer Verlag. 1972.
- [Fujisaki 93] H. Fujisaki, M. Ljungqvist, H. Murata. *Analysis and modelling of word accent and sentence intonation in Swedish*. Proceedings of the ICASSP'93. 1993.
- [Fujisaki 95] H. Fujisaki, S. Ohno. *An analysis and modelling of fundamental frequency contour of English utterances*. Proceedings of the EUROSPEECH'95. Madrid. str. 985–988. 1995.
- [Fujisaki 96a] H. Fujisaki, S. Ohno, O. Tomita. *On the levels of accentuation in spoken Japanese*. Proceedings of the ICSLP'96. Philadelphia. str. 634–637. 1996.

- [Fujisaki 96b] H. Fujisaki, S. Ohno. Prosodic *Parametrization of spoken Japanese based on a model of the generation process of F0 contours*. Proceedings of the ICSLP'96. Philadelphia. str. 2439–2442. 1996.
- [Furui 89] S. Furui. *Digital Speech Processing, Synthesis and Recognition*. Marcel Dekker. 1989.
- [Gabioud 94] B. Gabioud. *Articulatory models in speech synthesis. Fundamentals of Synthesis and Speech Recognition*. Uredil E. Keller. John Wiley & Sons. str. 215–230. 1994.
- [Gaved 93] M. Gaved. *Pronunciation and text normalisation in applied text-to-speech systems*. Proceedings of the EUROSPEECH'93. Berlin. str. 897–900. 1993.
- [Le Goff 96] B. Le Goff, C. Benoit. *A text-to-audiovisual-speech synthesiser for French*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Gold 69] N. Gold, L. Rabiner. *Parallel processing techniques for estimating pitch periods of speech in the time domain*. Journal of the Acoustical Society of America. 46(2). str. 442–448. 1969.
- [Grabnar 96] M. Grabnar. *Generiranje govora v polzaprti domeni vremenska napoved na WWW*. Diplomsko delo. Fakulteta za računalništvo in informatiko. Univerza v Ljubljani. 1996.
- [Grenc 95] B. Grenc. *Izgradnja baze difonov z označenimi osnovnimi periodami*. Diplomsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1995.
- [Grice 92] M. L. Grice. *The Intonation and Interrogation in Palermo Italian: Implications for Intonation Theory*. Doktorska disertacija. University College London. 1992.
- [Grice 96] M. L. Grice, M. Reyelt, R. Benz Müller, J. Meyer, A. Batliner. *Consistency in transcription and labelling of German intonation with GToBI*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Griffin 84] D. W. Griffin, J. S. Lim. *Signal estimation from modified short-time Fourier transform*. IEEE Transactions on Acoustics, Speech and Signal Processing. (32). str. 236–243. 1984.
- [Grigoriu 96] A. Grigoriu, J. P. Vonwiller, R. W. King. *An automatic intonation tone contour labelling and classification algorithm*. Proceedings of the ICASSP'94. 1994.
- [Gros 91] J. Gros, L. Gyergyek, F. Mihelič. *Predstavitev samoglasniških fonemov z značilkami iz razvoja govornih samoglasniških fonemov po Karhunenu in po Loeveju*. Elektrotehniški vestnik. (58). str. 88–92. 1991.
- [Gros 95] J. Gros, N. Pavešič, S. Dobrišek, M. Erpič, B. Grenc, A. Rakar, T. Šef, V. Vračar, F. Mihelič. *Sistem za sintezo slovenskega govora*. Zbornik ERK'95. (B). Portorož. str. 265–268. 1995.
- [Gros 96a] J. Gros, N. Pavešič, F. Mihelič, S. Dobrišek. *A text-to-speech system*

- for the Slovenian language. Proceedings of the 3rd Slovenian–German and 2nd SDRV Workshop on Speech and Image Understanding. Ljubljana. str. 47–57. 1996.
- [Gros 96b] J. Gros, S. Dobrišek, F. Mihelič, N. Pavešič. *Slovenian text-to-speech synthesis*. Proceedings of the MIPRO'96. str. 2.13–2.18. Opatija. 1996.
- [Gros 96c] J. Gros, I. Ipšič, S. Dobrišek, F. Mihelič, N. Pavešič. *Segmentation and labelling of Slovenian diphone inventories*. Proceedings of the COLING'96. Copenhagen. str. 298–303. 1996.
- [Gros 96d] J. Gros, S. Dobrišek, F. Mihelič, N. Pavešič. *A text-to-speech system for the Slovenian language*. Proceedings of the EUSIPCO'96. Trst. str. 1043–1046. 1996.
- [Gros 96e] J. Gros, N. Pavešič, F. Mihelič, M. Benedik. *Analiza časovnega trajanja govornih enot pri različnih hitrostih govora*. Zbornik ERK'96. Portorož. (B). str. 243–246. 1996.
- [Gros 97a] J. Gros, N. Pavešič, F. Mihelič. *Speech timing in Slovenian text-to-speech synthesis*. Proceedings of the 2nd SQEL Workshop on Multi-Lingual Information Retrieval Dialogs. Plzen. str. 137–140. 1997.
- [Gros 97b] J. Gros, N. Pavešič, F. Mihelič. *Text-to-speech synthesis: a complete system for the Slovenian language*. Journal of Computing and Information Technology. (CIT). V tisku.
- [Gros 97c] J. Gros, N. Pavešič, F. Mihelič. *Syllable and segment duration at different speaking rates in the Slovenian language*. Proceedings of the EUROSPEECH'97. Vol. 1. Rodos. 1997.
- [Gros 97d] J. Gros, N. Pavešič, F. Mihelič. *Duration modelling in Slovenian TTS*. Proceedings of the EUROSPEECH'97. Vol. 2. Rodos. 1997.
- [Gros 97e] J. Gros. *Samodejno tvorjenje govora iz besedil*. Doktorska disertacija. Fakulteta za elektrotehniko. Univerza v Ljubljani. 1997.
- [Gros 99] J. Gros. *Dvostopenjski model trajanja za slovenski jezik*. Elektrotehniški Vestnik. Let. 66. Št. 2. str. 92–97. 1999.
- [Grosjean 75] F. Grosjean, A. Deschamps. *Analyse contrastive des variables temporelles de l'anglais et du français*. *Phonetica*. (31). str. 144–184. 1975.
- [Gussenhoven 84] C. Gussenhoven. *On the Grammar and Semantics of Sentence Accents*. Floris Publications. Dordrecht. 1984.
- [Gyergyek 91] L. Gyergyek. *Teorija signalov in obdelava signalov*. Založba ZAFER. 1991.
- [Halliday 70] M. A. K. Halliday. *A Course in Spoken English: Intonation*. Oxford University Press. Oxford. 1970.
- [Hamon 88] C. Hamon. *Procédé et dispositif de synthèse de parole par addition*

- recouvrement de formes d'onde*. Patentni zahtevk. FR 2 6636 163 – A1. 1988.
- [Hamon 89] C. Hamon, E. Moulines, C. Charpentier. *A diphone synthesis system based on time-domain prosodic modifications of speech*. Proceedings of the ICASSP'89. str. 238–241. 1989.
- [Haumont 94] G. Caelen–Haumont. *Semantic and pragmatic prediction of prosodic structures*. Fundamentals of Speech Synthesis and Speech Recognition. Uredil E. Keller. John Wiley & Sons. str. 270–293. 1994.
- [Hirai 95] T. Hirai, N. Higuchi, Y. Sagisaka. *Automatic detection of major phrase boundaries using statistical properties of superpositional F0 control model parameters*. Proceedings of the EUROSPEECH'95. Madrid. str. 1341–1344. 1995.
- [Hirschberg 93] J. Hirschberg. *Pitch accent in context: predicting intonational prominence from text*. Artificial Intelligence. (63). str. 305–340. 1993.
- [Hirst 93] D. J. Hirst, R. Espresser. *Automatic modelling of fundamental frequency using a quadratic spline function*. Travaux de l'Institut de Phonétique d'Aix. (15). 1993.
- [Hirst 95] D. J. Hirst. *Prosodic labelling tools*. MULTEXT-LRE report. 1995.
- [Holmes 64] J. Holmes, Y. Mattingly, J. Shearme. *Speech synthesis by rule*. Language and speech. (7). str. 127–143. 1964.
- [Höge 94] H. Höge. *Speech technology for telephone applications*. Proceedings CRIM/FORWISS Workshop on Progress and Prospects of Speech Research and Technology. München. str. 10–1/10–16. 1994.
- [Horne 96] M. Horne, M. Filipsson. *Implementation and evaluation of a model for synthesis of Swedish intonation*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [House 96] D. House. *Differential perception of tonal contours through the syllable*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Hribar 84] J. Hribar. *Sinteza umetnega govora iz teksta*. Magistrsko delo. Univerza v Ljubljani. Fakulteta za elektrotehniko. 1984.
- [Huang 97] X. Huang, A. Acero, H. Hon, Y. Ju, J. Liu, S. Meredith, M. Plumpe. *Recent improvements on Microsoft's trainable text-to-speech system WHISTLER*. Proceedings of the ICASSP'97. München. 1997.
- [Huggins 71] A. W. F. Huggins. *Just noticeable differences for segment duration in natural speech*. Journal of the Acoustical Society of America. (51). str. 1270–1278. 1971.
- [Inkelas 88] S. Inkelas, D. Zec. *Serbo-Croatian pitch accent: the interaction of tone, stress and intonation*. Language. (64). str. 227–248. 1988.
- [IPA 89] *Report on the 1989 Kiel Convention*. Journal of the International Phonetic Association. (19). str. 67–80. 1989.
- [Ipšić 95] I. Ipšić, F. Mihelič, J. Gros, K. Pepelnjak, S. Dobrišek, N. Pavešić.

- An overview of the Spoken Queries in European Languages project.* Zbornik ERK'95. (B). str. 281–284. Portorož. 1995.
- [Ipšič 96] I. Ipšič. *Razpoznavanje besed v vezanem govoru.* Doktorska disertacija. Fakulteta za elektrotehniko. Univerza v Ljubljani. 1996.
- [Ipšič 97] I. Ipšič, F. Mihelič, K. Pepelnjak, J. Gros, S. Dobrišek, N. Pavešič, E. Nöth. *The Slovenian dialog system for air flight inquiries.* Proceedings of the 2nd SQEL Workshop on Multi-Lingual Information Retrieval Dialogs. Plzen. str. 133–136. 1997.
- [ITU 88] ITU–T. *Recommendation P.80. Methods for subjective determination of transmission quality.* Blue Book. (5). str. 197–198. 1988.
- [ITU 93] ITU–T. *Recommendation P.81. Telephone transmission quality subjective opinion tests – Modulated noise reference unit.* Blue Book. (5). str. 1–5. 1993.
- [ITU 94] ITU–T. *Recommendation P.85. Telephone transmission quality subjective tests – a method for subjective performance assessment of the quality of speech voice output devices.* Blue Book. (5). str. 1–9. 1994.
- [Jekosch 93] U. Jekosch. *Speech quality assessment.* Proceedings of the EUROSPEECH'93. Berlin. str. 1387–1394. 1993.
- [Jensen 93] U. Jensen, R.K. Moore, P. Dalsgaard, B. Lindberg. *Modelling of intonation contours at the sentence level using CHMMs and the 1961 O'Connor and Arnold scheme.* Proceedings of the EUROSPEECH'93. Berlin. str. 785–788. 1993.
- [Kaeslin 86] H. Kaeslin. *A systematic approach to the extraction of diphone elements from natural speech.* IEEE Transactions on Acoustics, Speech and Signal Processing. (ASSP–34). str. 264–271. 1986.
- [Kang 85] G.S. Kang, S. Everett. *Improvement of the excitation source in the narrow-band linear prediction vocoder.* IEEE Transactions on Acoustics, Speech and Signal Processing. (ASSP–33). str. 377–386. 1985.
- [Katae 96] N. Katae, S. Kimura. *Natural prosody generation for domain specific text-to-speech systems.* Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Karn 96] H. E. Karn. *Design and evaluation of a phonological phrase parser for Spanish text-to-speech.* Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Keller 94] E. Keller. *Fundamentals of Speech Synthesis and Speech Recognition.* John Wiley & Sons. 1994.
- [Klatt 73] D. H. Klatt. *Interaction between two factors that influence vowel duration.* Journal of the Acoustical Society of America. (54). str. 1102–1104. 1973.
- [Klatt 76a] D. H. Klatt. *Structure of a phonological rule component for a synthe-*

- sis by rule program*. IEEE Transactions on Acoustics, Speech and Signal Processing (ASSP-24). str. 391–398. 1976.
- [Klatt 76b] D.H. Klatt. *Linguistic uses of segmental duration in English: acoustic and perceptual evidence*. Journal of the Acoustical Society of America. (59). str. 1209–1221. 1976.
- [Klatt 87] D. H. Klatt. *Review of text-to-speech conversion for English*. Journal of the Acoustical Society of America. 82(3). str. 737–793. 1987.
- [Klaus 95] H. Klaus, A. Niebank. *Comparison of reference system approaches for the quality assessment of synthesised speech*. Proceedings of the EUROSSPEECH'95. Madrid. str. 1267–1269. 1995.
- [Komar 96] S. Komar. *Funkcija intonacije v diskurzivni analizi intervjuja v angleščini in slovenščini*. Doktorska disertacija. Filozofska fakulteta. Univerza v Ljubljani. 1996.
- [Ladd 83] D. R. Ladd. *Phonological features of intonation peaks*. Language. (59). str. 721–759. 1983.
- [Laroche 93] J. Laroche, Y. Stylianou, E. Moulines. *HNS: Speech modification based on a harmonic + noise model*. Proceedings of the ICASSP'93. str. 550–553. 1993.
- [Lehiste 72] I. Lehiste. *The timing of utterances and linguistic boundaries*. Journal of the Acoustical Society of America. (51). str. 2018–2024. 1987.
- [Lehiste 77] I. Lehiste. *Isochrony reconsidered*. Journal of Phonetics. (5). str. 253–263. 1977.
- [Lieberman 89] M. Y. Liberman, K. W. Church. *Text analysis and word pronunciation in text-to-speech synthesis*. Proceedings of the DARPA Workshop on Speech and Natural Language Processing. 1989.
- [Lieberman 92] M. Y. Liberman, K. W. Church. *Text analysis and word pronunciation in text-to-speech synthesis*. Advances in Speech Signal Processing. Uredila S. Furui, M. M. Sondhi. Marcel Dekker. New York. str. 791–831. 1992.
- [Lindblom 90] B. Lindblom. *Explaining phonetic variation*. Speech Production and Speech Modelling. Uredila W.J. Hardcastle, A. Marchal. Kluwer. Dordrecht. 1990.
- [Lindstrom 93] A. Lindstrom, M. Ljungqvist, K. Gustafson. *A modular architecture supporting multiple hypotheses for conversion of text to phonetic and linguistic entities*. Proceedings of the EUROSPEECH'93. Berlin. str. 1463–1467. 1993.
- [Ljolje 93] A. Ljolje, M. D. Riley. *Automatic segmentation of speech for TTS*. Proceedings of the EUROSPEECH'93. Berlin. str. 1445–1448. 1993.
- [Ljungqvist 93] M. Ljungqvist, H. Fujisaki. *Generating intonation for Swedish text-to-speech conversion using a quantitative model for the F0 con-*

- tour. Proceedings of the EUROSPEECH'93. Berlin. str. 873–876. 1993.
- [Llisteri 94] J. Llisteri, D. Poch–Olive. *Phonetics and phonology of speaking styles*. Specifications and Standards. MULTEXT–LRE projekt. 1994.
- [López–Gonzalo 96] E. López–Gonzalo. *Statistical methods in data-driven modeling of Spanish prosody for text to speech*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Lucassen 84] J. M. Lucassen, R. L. Mercer. *An information theoretic approach to the automatic determination of phonemic baseforms*. Proceedings of the ICASSP'92. San Francisco. 1984.
- [Mazej 97] U. Mazej. *Načrtovanje preizkusa za ocenjevanje kakovosti sintetizirane govora*. Diplomsko delo. Fakulteta za elektrotehniko. Univerza v Ljubljani. 1997.
- [McCulloch 87] N. McCulloch, M. Bedworth, J. Bridle. *NETSPEAK – a reimplementation of NETTALK*. Computer Speech and Language. str. 289–301. 1987.
- [Medan 91] Y. Medan, E. Yair, D. Chazan. *Super resolution pitch determination of speech signals*. IEEE Transactions on Signal Processing. 39(1). str. 40–48. 1991.
- [Merten 87] P. Merten. *L'intonation du Français: de la description linguistique à la reconnaissance automatique*. Doktorska disertacija. Katoliška univerza v Leuvnu. 1987.
- [Meyer 93] P. Meyer, H. W. Ruhl, R. Kruger, M. Kugler. *PHRITTS – a text-to-speech synthesiser for the German language*. Proceedings of the EUROSPEECH'93. Berlin. str. 877–880. 1993.
- [Mihelič 91] F. Mihelič. *Akustično fonetična pretvorba slovenskega govora*. Doktorska disertacija. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1991.
- [Mihelič 93a] F. Mihelič, B. Lozej. *Trajanje in glasnost slovenskih fonemov*, Zbornik druge elektrotehniške in računalniške konference ERK'93. str. 267–270. 1993.
- [Mihelič 93b] F. Mihelič, B. Lozej. *Dolžine trajanja in glasnost izgovarjave slovenskih fonemov*. Zbornik Jezik tako in drugače. Ljubljana. str. 441–451. 1993.
- [Mihelič 94] F. Mihelič, S. Dobrišek, J. Gros, I. Ipšič, K. Pepelnjak, N. Pavešič. *Development of a Continuous Speech Recognition System for Information Services*. Proceedings of the International Workshop on Modern Modes of Man–Machine Communication. Maribor. str. 17.1–17.20. 1994.
- [Miki 93] N. Miki, N. Kamiyama, N. Nagai. *Synthesis and analysis of vocal*

- source with vibration of larynx*. Proceedings of the EURO-SPEECH'93. Berlin. str. 421–424. 1993.
- [Mixdorff 95] H. Mixdorff, H. Fujisaki. *A scheme for a model-based synthesis by rule of F0 contours of German utterances*. Proceedings of the EUROSPEECH'95. Madrid. str. 1823–1826. 1995.
- [Miksič 93] A. Miksič, Z. Kačič. *Analiza ritma glasbenega in govornega signala*. Elektrotehniški vestnik. Ljubljana. 60(1). str. 55–58. 1993.
- [Minematsu 96] N. Minematsu, S. Nakagawa, K. Hirose. *Prosodic manipulation system of speech material for perceptual experiments*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Möbius 96] B. Möbius, J. van Santen. *Modeling segmental duration in German text-to-speech synthesis*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Morton 93] S. J. Young, F. Fallside. *Speech synthesis from context*. Journal of the Acoustical Society of America. (66). str. 685–695. 1979.
- [Moulines 90] E. Moulines, F. Charpentier. *Pitch-Synchronous waveform processing techniques for text-to-speech synthesis using diphones*. Speech Communication. (9). 453–467. 1990.
- [Muha 92] A. Vidovič-Muha. *Slovenski jezik I*. Zapiski s predavanj. 1992/93.
- [Mumolo 96] E. Mumolo, M. Teia. *Prosody generation by means of a syntactic approach and its application in a text-to-speech system*. Proceedings of the EUSIPCO'96. Trst. 1996.
- [Natvig 94] J. E. Natvig. *Study and prediction of syllable duration in read aloud Norwegian – preliminary results*. COST–233 Meeting. Lausanne. 1994.
- [Neweklowsky 73] G. Neweklowsky. *Spezifische Dauer und Tonhöhe der Vokale*. Phonetica. (32). str. 38–60. 1973.
- [Newman 84] W. M. Newman, R. Sproul. *Principles of Interactive Computer Graphics*. McGraw–Hill. 1984.
- [Ni 96] J. Ni, R. Wang, D. Xia. *A functional model for generation of the local components of F0 contours in Chinese*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Nöth 90] E. Nöth. *Prosodische Information in der Automatischen Spracherkennung Berechnung und Anwendung*. Doktorska disertacija. Univerza Erlangen – Nürnberg. 1990.
- [Noll 67] A. M. Noll. *Cepstrum pitch determination*. Journal of the Acoustical Society of America. 41(2). str. 293–309. 1967.
- [O'Connor 73] J. D. O'Connor, G.R. Arnold. *Intonation of Colloquial English*. Longman. London. 1973.
- [Olaszy 91] G. Olaszy. *Adaptation of the MULTIVOX TTS system to Italian*. Proceedings of the EUROSPEECH'91. Genova. str. 1247–1250. 1991.

- [Omerza 70] Z. Omerza. *Uporabna fonetika*. Državna založba Slovenije. Ljubljana. 1970.
- [Orešnik 84] J. Orešnik. *Slovenske breznaglasnice se vedejo predvsem kot proklitike*. Jezik in slovstvo. (4). 1983/84.
- [Ottesen 93] G. E. Ottesen. *An Automatic Diphone Segmentation System*. SINTEF DELAB N-7034. Proceedings of the EUROSPEECH'91. Genova. str. 713-716. 1993.
- [Pavešič 81] N. Pavešič, F. Mihelič. *Statistične metode*. Fakulteta za elektrotehniko v Ljubljani. 1981.
- [Pavešič 92] N. Pavešič. *Razpoznavanje vzorcev*. Univerza v Ljubljani. Fakulteta za elektrotehniko in računalništvo. 1992.
- [Pavešič 96] N. Pavešič, J. Gros, S. Dobrišek, F. Mihelič. *Slovenian text-to-speech synthesis*. Proceedings of the SOFTCOM'96. Split. str. 99-110. 1996.
- [Penrose 89] R. Penrose. *The Emperor's New Mind*. Vintage. Oxford University Press. 1989.
- [Petek 96] B. Petek, R. Šuštaršič, S. Komar. *An acoustic analysis of contemporary vowels of the Standard Slovenian language*. Proceedings of the ICSLP'96. (1). Philadelphia. str. 133-136. 1996.
- [Peterson 60] G. E. Peterson, I. Lehiste. *Duration of syllable nuclei in English*. Journal of the Acoustical Society of America. (32). str. 693-703. 1960.
- [Pfister 94] B. Pfister, C. Traber. *Text-to-speech synthesis: an introduction and a case study*. *Fundamentals of Synthesis and Speech Recognition*. Uredil E. Keller. John Wiley & Sons. str. 87-107. 1994.
- [Phillips 85] M. S. Phillips. *A feature-based time domain pitch tracker*. Journal of the Acoustical Society of America. (77). str. S9-S210(A). 1985.
- [Pierrehumbert 88] J. B. Pierrehumbert, M.E. Beckman. *Japanese Tone Structure*. The MIT Press. Cambridge, Massachusetts. 1988.
- [Pierrel 87] J. M. Pierrel. *Dialogue Homme-Machine*. Hermes. Paris. 1987.
- [Pols 89] L. C. W. Pols. *Assessment of text-to-speech synthesis systems*. *Speech Input and Output Assessment – Multilingual Methods and Standards*. Uredili A. Fourcin, G. Harland, W. Barry, V. Hazan. Ellis Horwood Ltd. John Willey & Sons. New York – Chichester – Brisbane – Toronto. str. 53-81. 1989.
- [Pols 94] L. C. W. Pols. *Synthesis performance assessment*. Proceedings of the CRIM/FORWISS Workshop on Progress and Prospects of Speech Research and Technology. Uredili H. Niemann in drugi. München. 1994.
- [Pols 95] L. C. W. Pols. *Speech synthesis evaluation*. *Survey of the State of the Art in Human Language Technology*. Uredil R.A. Cole. str. 500-501. 1995.

- [Portele 94] T. Portele, B. Heuft, F. Hofer, H. Meyer, W. Hess. *A new high quality speech synthesis system for German*. Proceedings of the CRIM/FORWISS Workshop on Progress and Prospects of Speech Research and Technology. Uredili H. Niemann in drugi. München. str. 274–277. 1994.
- [Portele 96] T. Portele, K.H. Stober, H. Meyer, W. Hess. *Generation of multiple synthesis inventories by a bootstrapping procedure*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Rabiner 89] L. Rabiner. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE. 77(2). str. 257–289. 1989.
- [Rakar 95] A. Rakar. *Modul makroprozodike za oblikovanje stavčne intonacije v okviru sinteze slovenskega govora*. Diplomsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1995.
- [Ried i95] M. Riedi. *A neural-network-based model of segmental duration for speech synthesis*. Proceedings of the EUROSPEECH'95. Madrid. 1995.
- [Rigler 68] J. Rigler. *Problematika naglaševanja v slovenskem knjižnem jeziku*. Jezik in slovstvo. (13). str. 192–199. 1968.
- [Riley 94] M. Riley, R. Sproat. *Text analysis tools in spoken language processing*. Proceedings of the 32nd Annual Meeting of the ACL. Las Cruces. 1994.
- [Roelands 93] M. Roelands, W. Verhelst. *Waveform similarity based overlap-add (WSOLA) for time-scale modification of speech: structures and evaluation*. Proceedings of the EUROSPEECH'93. Berlin. str. 337–340. 1993.
- [Ross 74] M. J. Ross, H. L. Schaffer, A. Cohen, R. Freudberg, H. J. Manley. *Average magnitude difference function pitch extractor*. IEEE Transactions on Acoustics, Speech and Signal Processing. (ASSP–25). str. 565–571. 1975.
- [Sagisaka 92] Y. Sagisaka, N. Kaiki, N. Iwahashi, K. Mimura. *ATR n-talk speech synthesis system*. Proceedings of the ICSLP'92. str. 483–486. 1992.
- [van Santen 90] J. P. H. van Santen, J. Olive. *The analysis of contextual effects on segmental duration*. Computer Speech & Language. (4). str. 365–390. 1990.
- [van Santen 92] J. P. H. van Santen. *Contextual effects on vowel duration*. Speech Communication. (11). str. 513–546. 1992.
- [van Santen 93] J. P. H. van Santen. *Timing in text-to-speech systems*. Proceedings of the EUROSPEECH'93. Berlin. str. 3397–1400. 1993.
- [SAZU 90] Slovenska akademija znanosti in umetnosti (SAZU). *Slovenski pravopis 1: Pravila*. Državna založba Slovenije. Ljubljana. 1990.

- [Scheffers 88] M. T. M. Scheffers. *Automatic stylization of F0 contours*. Proceedings of the 5th FASE Symposium. str. 981–987. 1988.
- [Schmidt 93] M. Schmidt, S. Fitt, C. Scott, M. Jack. *Phonetic transcription standards for European names (ONOMASTICA)*. Proceedings of the EUROSPEECH'93. Berlin. str. 279–280. 1993.
- [Schroeder 68] M. R. Schroeder. *Period histogram and product spectrum: New methods for fundamental frequency measurement*. Journal of the Acoustical Society of America. 43(4). str. 829–834. 1968.
- [Scully 87] C. Scully. *Linguistic units and units of speech production*. *Speech Communication*. (6). str. 77–142. 1987.
- [Senčar 96] T. Senčar. *Samodejno pridobivanje nabora slovenskih difonov: določanje mej med difoni in znotraj difonov*. Diplomsko delo. Fakulteta za elektrotehniko. Univerza v Ljubljani. 1996.
- [Shaughnessy 77] D. O'Shaughnessy. *Fundamental frequency by rule for a text-to-speech system*. Proceedings of the ICASSP'77. New York. str. 571–574. 1977.
- [Shaughnessy 84] D. O'Shaughnessy. *Design of a real-time French text-to-speech synthesiser*. *Speech Communication*. (3). str. 233–243. 1984.
- [Shaughnessy 88] D. O'Shaughnessy, L. Barbeau, D. Bernardi, D. Archambault. *Diphone speech synthesis*. *Speech Communication*. (7). str. 55–65. 1988.
- [Shaughnessy 89] D. O'Shaughnessy. *Parsing with a small dictionary for applications such as text-to-speech*. *Computational Linguistics*. (15). str. 97–108. 1989.
- [Shaughnessy 95] D. O'Shaughnessy. *Timing patterns in fluent and disfluent spontaneous speech*. Proceedings of the ICASSP'95. Adelaide. str. 600–603. 1995.
- [Shimodaira 93] H. Shimodaira, M. Nakai. *Accent phrase segmentation using transition probabilities between pitch pattern templates*. Proceedings of the EUROSPEECH'93. Berlin. str. 1767–1770. 1993.
- [Sonntag 97] G. P. Sonntag, T. Portele, B. Heuft. *Prosody generation with a neural network: weighing the importance of input parameters*. Proceedings of the ICASSP'97. München. 1997.
- [van Son 93] R. J. J. H. van Son, L. C. W. Pols. *How does speaking rate influence vowel tract parameters? Analysis and Synthesis of Speech*. Mouton de Gruyter. Berlin – New York. 1993.
- [Sorin 87] C. Sorin, D. Larreur, R. Llorca. *A rhythm based prosodic parser for text-to-speech systems in French*. Proceedings of the 11th ICP-hSc. Tallin. Estonija. str. 125–128. 1987.
- [Sorin 94] C. Sorin. *Towards High-Quality Multilingual Text-to-Speech*. Proceedings of the CRIM/FORWISS Workshop on Progress and Pros-

- pects of Speech Research and Technology. Uredili H. Niemann in drugi. München. str. 53–62. 1994.
- [Sovdat 96] M. Sovdat. *Samodejno pridobivanje nabora slovenskih difonov: določanje osnovne periode*. Diplomsko delo. Fakulteta za elektrotehniko. Univerza v Ljubljani. 1996.
- [Sproat 95] R. Sproat. *Text interpretation for TtS synthesis. Survey of the State of the Art in Human Language Technology*. Uredil R.A. Cole str. 202–209. 1995.
- [Srebot 75a] T. Srebot Rejec. *Soglasniški sklopi v slovenščini in kontrastivna analiza angleških in slovenskih soglasniških sklopov*. Magistrsko delo. Filozofska fakulteta. Univerza v Ljubljani. 1975.
- [Srebot 75b] T. Srebot Rejec. *Začetni in končni soglasniški sklopi v slovenskem knjižnem jeziku*. Slavistična revija. (23). str. 289–320. 1975.
- [Srebot 81] T. Srebot Rejec. *On the allophones of /v/ in standard Slovene*. Scandolinguistica. (27). str. 233–241. 1981.
- [Srebot 87] T. Srebot Rejec. *The sound systems of English and Slovene compared: a distinctive feature analysis*. Linguistica. (27). 1987.
- [Srebot 88] T. Srebot Rejec. *Word Accent and Vowel Duration in Standard Slovene: An Acoustic and Linguistic Investigation*. Slavistische Beiträge. Band 226. Verlag Otto Sagner. München. 1988.
- [Srebot 89] T. Srebot Rejec. *Kakovost slovenskih in angleških samoglasnikov*. Jezik in slovstvo. (34). str. 57–64. 1988–89.
- [Srebot 92] T. Srebot Rejec. *Initial and final sonorant clusters in Slovene*. Linguistica. Ljubljana. (32). str. 227–230. 1992.
- [Srebot 96] T. Srebot Rejec. *Memorandum in matematika v slovenščini in angleščini*. SAZU. Razred za filološke in literarne vede. Razprave. (15). str. 155–160. 1996.
- [Schukat 92] E. G. Schukat–Talamazzini et al. *Acoustic modelling of subword units in the ISADORA speech recognizer*. Proceedings of the ICASSP'92. San Francisco. str. 577–580. 1992.
- [SSKJ 94] *Slovar slovenskega knjižnega jezika*. Državna založba Slovenije. 1994.
- [Styger 94] T. Styger, E. Keller. *Formant synthesis. Fundamentals of Synthesis and Speech Recognition*. Uredil E. Keller. John Wiley & Sons. str. 109–128. 1994.
- [Šef 95] T. Šef. *Spreminjanje prozodičnih lastnosti osnovnih enot sintetiziranega govora*. Diplomsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1995.
- [Šef 96] T. Šef, A. Dobnikar. *Grafemsko-fonemski modul za sintezo slovenskega govora*. Zbornik ERK'96. (B). Portorož. str. 227–230. 1996.
- [Šef 98] T. Šef. *Sistem za govorno posredovanje obvestil o prostih delov-*

- nih mestih*. Magistrsko delo. Fakulteta za računalništvo in informatiko. Univerza v Ljubljani. 1995.
- [Šuštaršič 94] R. Šuštaršič. *Kontrastivna analiza angleške in slovenske stavčne intonacije*. Doktorska disertacija. Filozofska fakulteta. Univerza v Ljubljani. 1994.
- [t'Hart 73] J. t'Hart, A. Cohen. *Intonation by rule: a perceptual quest*. Journal of phonetics. (1). str. 309–327. 1973.
- [Taylor 91] P. A. Taylor and S. D. Isard. *Automatic diphone segmentation*. Proceedings of the EUROSPEECH'91. Genova. str. 709–712. 1991.
- [Taylor 92] P. A. Taylor, S. D. Isard. *A new model of intonation for use with speech synthesis and recognition*. Proceedings of the ICSLP'92. Banff. 1992.
- [Taylor 93] P. Taylor. *Automatic recognition of intonation from F0 contours using the rise fall connection model*. Proceedings of the EUROSPEECH'93. Berlin. str. 789–792. 1993.
- [Tench 96] P. Tench. *The Intonation Systems of English*. Založba Cassel. 1996.
- [Tičević 00] E. Tičević. *Samodejna analiza naglašenosti večzložnih besed*. Diplomsko delo. Fakulteta za elektrotehniko. Univerza v Ljubljani. 2000.
- [Toporišič 67] J. Toporišič. *Stavčna intonacija*. Jezikovni pogovori II. Cankarjeva založba. Ljubljana. str. 167–172. 1967.
- [Toporišič 68] J. Toporišič. *Liki slovenskih tonemov*. Slavistična revija. (16). str. 315–393. 1968.
- [Toporišič 69] J. Toporišič. *Slovenski jezik, izgovor i intonacija s recitacijama na pločama*. 1969.
- [Toporišič 78] J. Toporišič. *Glasovna in naglasna podoba slovenskega jezika*. Založba Obzorja Maribor. 1987.
- [Toporišič 91] J. Toporišič. *Slovenska slovnica*. Založba Obzorja. Maribor. 1991.
- [Toporišič 92] J. Toporišič. *Enciklopedija slovenskega jezika*. Cankarjeva založba. Ljubljana. 1992.
- [Traber 93] C. Traber. *Syntactic processing and prosody control in the SVOX TTS system for German*. Proceedings of the EUROSPEECH'93. Berlin. str. 2099–3002. 1993.
- [Tzoukerman 95] E. Tzoukermann, O. Soumoy. *Segmental Duration in French Text-to-Speech Synthesis*. Proceedings of the EUROSPEECH'96. Madrid. str. 607–610. 1995.
- [Vitez 95] P. Vitez, V. Aubergé. *Intonation gesture of Slovene: first indications*. Proceedings of the EUROSPEECH'95. Madrid. str. 2073–2075. 1995.
- [Vogel 96] I. Vogel, S. Hoskins. *On the interaction of clash, focus and phonological phrasing*. Proceedings of the ICSLP'96. Philadelphia. 1996.
- [Vračar 95] V. Vračar. *Modul za nastavljanje mikroprozodike vezanega bese-*

- dila v okviru sinteze slovenskega jezika. Diplomsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1995.
- [Vroomen 93] J. Vroomen, R. Collier, S. Mozzionacci. *Duration and intonation in emotional speech*. Proc. of the EUROSPEECH'93. Berlin. str. 577–580. 1993.
- [Weilguny 93] S. Weilguny. *Grafemsko-fonemski modul za sintezo izoliranih besed slovenskega jezika*. Magistrsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1993.
- [Wells 89] J. Wells. *Transcription labelling and reference. peech Input and Output Assessment – Multilingual Methods and Standards*. Uredili A. Fourcin, G. Harland, W. Barry, V. Hazan. Ellis Horwood Ltd. John Willey & Sons. New York – Chichester – Brisbane – Toronto. str. 141–187. 1989.
- [Wells 92] J. Wells, W. Barry, M. Grice, A. Fourcin, D. Gibbon. *Standard Computer-Compatible Transcription*. ESPRIT Project 2589 (SAM). SAM-UCL-037. 1992.
- [Werner 94] S. Werner, E. Keller. *Prosodic aspects of speech*. Fundamentals of Speech Synthesis and Speech Recognition. Uredil E. Keller. John Wiley & Sons. str. 24–39. 1994.
- [Yamashita 94] Y. Yamashita, R. Mizoguchi. *Automatic generation of prosodic rules for speech synthesis*. Proceedings of the ICASSP'94. 1994.
- [Young 79] S. J. Young, F. Fallside. *Speech synthesis from context*. Journal of the American Acoustical Society. (66). str. 685–695. 1979.
- [Zellner 94] B. Zellner. *The temporal structures of speech*. Fundamentals of Speech Synthesis and Speech Recognition. Uredil E. Keller. John Wiley & Sons. str. 40–62. 1994.
- [Žagar 87] F. Žagar. *Slovenska slovnica in jezikovna vadnica*. Založba Obzorja Maribor. 1987.

Tu navajamo razlage nekaterih izrazov, ki so pomembni za razumevanje dela. Ker nam je inženirjem jezikoslovno izrazje manj znano, pomensko opisujemo predvsem jezikoslovne izraze. Razlage večine izrazov smo neposredno povzeli po [SAZU 90, str. 199–212] in [Toporišič 92].

akcent	gl. naglas
akut	1. tonemski naglas; uresničuje se kot nizki ton na naglašnem zlogu, ki mu sledi višji ton na ponaglašnem zlogu, če le-ta obstaja 2. naglasno znamenje /' / oziroma /ˈ /; gl. ostrivec, krativec
alofon	konkretna pojavna oblika fonema z določeno skupno značilnostjo na izgovorni ali slušni ravni; navadno nastopa v posebnem glasovnem okolju: npr. [F] namesto [m] pred <f> ali <v>; sopom. fonemska varianta ali fonemska različica
antikadenca	zadnji del končnega člena stavčne intonacije, nasproten zaključku pripovednih povedi, pri katerem gremo z glasom navzgor
asimilacija	gl. prilikovanje
bariton	jezikoslovni izraz za večzložno besedo, ki nima naglasa na zadnjem zlogu
besedni naglas	gl. naglas; sopom. leksikalni naglas
breznaglasnica	1. beseda, ki nima naglasa; sopom. naslonka ali klitika 2. beseda, ki v določeni zvezi nima naglasa; npr. besedi dober in dobro v pozdravih <Dober dan.> <Dobro jutro.>
cifra	gl. številka
cirkumfleks	1. tonemski naglas; uresničuje se kot visoki ton na naglašnem zlogu in nižji ton na ponaglašnem zlogu, če ta obstaja 2. naglasno znamenje, gl. dvojni gravis
členitev s premori	enostavno poved, zlasti pa večstavčno, v govoru navadno členimo s premori; tipična mesta členitve so: meja med stavki iste povedi (zlasti v priredju), med vrivkom, pristavkom, dostavkom, izpostavkom, pastavkom in preo-

členitev z intonacijo	<p>stalim delom povedi ali stavka. Vsi ti premori imajo različne, a tipične dolžine.</p> <p>govorjeno besedilo členimo ne le s premori, ampak tudi z intonacijo; navadno ima vsak s premori ločen segment intonacijo s tipičnim intonacijskim zaključkom, s t. i. intonacijsko glavo (samo segmenti, nastali s premori zaradi počasnega govorjenja jih navadno nimajo), tako da jih več tvori intonacijsko enoto. Pri hitrem govoru se členitev s premori lahko izgubi, ostaja pa členitev z intonacijo, deljena intonacija pa se strne v enoto.</p>
črkovalna kratica	<p>poimenovanje, navadno iz samih začetnic besedne zveze, npr. <SZDL>, ki jo beremo [s@z@d@l@], torej z imeni črk, ne pa z njihovo navadno glasovno vrednostjo, za razliko od nečrkovalne kratice <TOMOS></p>
diakritično znamenje dinamični naglas difon	<p>ločevalno, naglasno znamenje</p> <p>gl. jakostni naglas</p> <p>govorna enota, ki zajema glasovni prehod, tj. zadnji del glasu in prvi del naslednjega; pred premorom se izhodni del glasu razteza</p>
dolgi soglasnik	<p>soglasnik, ki se piše z dvema črkama, izjemoma tudi s tremi; prim. <i>izza</i>, <i>vas s stolpom</i></p>
drsnik dvoglasnik	<p>zvočnika [v] ali [j]; prvi je temen, drugi svetel</p> <p>1. zveza dveh glasovnih prvin z vrednostjo enega fonema;</p> <p>2. zveza samoglasnika in zvočnika, zlasti obeh drsnikov; prim. <daj>, <prav>; sopom. diftong</p>
dvoglasniški [w]	<p>varianta fonema /v/ v položaju za samoglasnikom; prim. <prav> [praw]</p>
dvojnični naglas	<p>lastnost, da ima beseda ali njena oblika več naglasnih možnosti:</p> <p>1. na različnih zlogih, npr. <krožite> [kro:ZitE] in [kr-OZi:tE]</p> <p>2. da je naglašena ali ne, npr. <ampak> [ampak] in [a:mpak]</p>
dvojni gravis elipsa enakozložnica enakozvočnica	<p>znamenje za cirkumfleks na kratkem zlogu, npr. br`at</p> <p>gl. izpust</p> <p>beseda, ki ima toliko zlogov kot katera druga</p> <p>beseda, ki se enako piše (enakopisnica) ali izgovarja (enakoglasnica) kot katera druga</p>
enklitika	<p>1. gl. naslonka</p> <p>2. beseda, ki se naglasno opira na naglašeno besedo pred njo; npr. <Ubranili so se ga.></p>

enostavčna poved	poved, ki sestoji iz enega samega stavka: navadnega glagolskega <Danes je lepo vreme.>, polstavčnega <Pisoč pismo.> ali pastavčnega <Joj!>
enosegmentna poved	poved, ki se izgovarja brez vsakega vmesnega premora
enotna izreka	pravorečna določitev, ki velja za vse, ki govorijo knjižno
favkalni soglasnik	zobnovenčni ali ustnični zapornik, izgovarjan z odporo skozi nos, npr. v besedah ali besednih zvezah dno , tnalo , obme
fon	gl. glas
fonem	glas ali skupina glasovnih različic, gl. alofon, ki nasproti drugim glasovom oz. alofonom v danem jeziku razločuje pomen besed ali morfemov, kot npr. /p/ in /v/ v paru /pas/ in /vas/; slovenski knjižni jezik pozna 29 fonemov [Toporišič 92, str. 41]
fonemska različica	gl. alofon
fonetika	jezikoslovna veda, ki proučuje tvorno plat jezika od glasu in naglasa do besedila; osnovna fonetična enota je glas ali fon. Govorimo lahko o fonetiki glasov, zloga, morfema, besede, besedne zveze, stavka, povedi, besedila. Svoje enote zaznamuje z oglatim oklepajem in pokončno pisavo črk [].
fonologija	del glasoslovja, ki proučuje pomenskorazločevalno vlogo pojavov zvočne strani jezika; njene osnovne enote so fonem, tonem, silabem ipd. Svoje enote zapisuje v poševnih oklepajih / /.
fonološki sestavi	poleg glasovnih sestavov, poznamo še prozodične, intonemske, registrske, hitrostne ipd. 1. prozodični fonološki sestavi so jakostni, naglasnostni ipd 2. intonemske enote so končne ali nekončne, padajoče oz. nepadajoče, nepadajoče pa so stopničaste, rastoče ali lebdeče 3. izrazitostne enote so naglašene ali nenaglašene 4. registrske enote so srednje ali nesrednje, nesrednje so visoke ali nizke, visoke pa so navadne ali falzetne 5. hitrostne enote so srednje ali nesrednje, nesrednje se nadalje delijo na upočasnjene ali pospešene
glas	zvok, tvorjen z govorili; akustična realizacija fonema; sopom. fon
glasovna vrednost črk	črke navadno zaznamujejo en alofon, vendar lahko ena črka v različnih zvezah zaznamuje tudi več alofonov; prim. črka <v> zaznamuje glasove [v], [w] in [W];

glava intonacije	del intonacijskega člena od vključno težišča naprej; lahko je končna (kadenčna ali antikadenčna) ali nekončna (polkadenčna); sopom. intonacijska glava
glavni naglas	desni naglas od dveh v besedi; zlasti če je prvi, navadno celo krepkejši, na predponi sestavljenke; v IPA zapisu ga zaznamujemo s [']
govorila	organi, ki jih uporabljamo pri tvorjenju glasov, prozodičnih in stavčnofonetičnih značilnosti; to so ustnice, sekalci, dlesni, trdo in mehko nebo z jezičkom, goltna stena, jezik, poklopec, nos in grlo, sapnik, sapnici, pljuča, prsni koš in prepona; sopom. govorni organi
govorjeni jezik grafemi	slušno uresničeno besedilo pisna znamenja za alofone, foneme ali pojave stavčne fonetike: črke, ločila, ideogrami ipd.
gravis	gl. krativec
hitri govor	pospešeno govorjenje, v katerem prihaja do izgovornih poenostavitvev, tudi izpada glasov
hitrost govora	tvorjenje različnega števila glasov na določeno časovno enoto
ideogram	pisno znamenje za pomen; prim. <%> (odstotek), <+> (plus)
inherentna lastnost	lastnost, ki ni odtujljiva; npr. inherentno trajanje trajanja glasu, inherentna osnovna frekvenca
intonacija	potek osnovnega tona v stavku; sestoji iz intonacijskega telesa in glave, iz enega ali več členov: če iz več, potem iz enega nekončnega (polkadenca) in enega končnega (kadenca ali antikadenca)
intonacijski člen	samostojni del intonacije kakšne povedi
intonacijski trup	gl. trup intonacije
intonacijsko črtovje	štiri vodoravne črte, katerih zgornji medprostor označuje intonacijsko višino 3 (težišče intonacije v dopolnjevalnih vprašanjih), srednji 2 (težišče intonacije v odločevalnih vprašanjih), spodnji 1 (iztek povedne intonacije); prostor nad črtovjem je za nenormalno zvišano, pod črtovjem pa za nenormalno znižano lego.
intonacijsko ogrodje	gl. intonacija
intonacijsko težišče	gl. težišče intonacije
intonem	intonacija v svoji pomenskorazločevalni vlogi: intonem 1 je intonacija, ki se končuje s kadenco (povedna intonacija), intonem 2 se končuje z antikadenco (vprašalna intonacija), intonem 3 je čustveno zaznamovan (vzklična intonacija, intonacija s poudarjanjem), ima pa bodisi po-

	večan ali zmanjšan interval na določeni tonski višini, ki je lahko tudi premaknjena z normale
IPA	mednarodna fonetična abeceda; angl. International Phonetic Alphabet
istoglasen	izgovarjan z istim glasom, npr. črka <d> v besedi <sladka> in <t> v besedi <potka>
izgovor izgovorljivi soglasniški sklop	slušna uresničitev glasovne strani jezika, zlasti knjižnega vsak soglasniški sklop, ki znotraj danega jezika ni naveden kot težko izgovorljiv ali neizgovorljiv
izpad glasu izpust	v govoru neuresničeni glas; sopom. elipsa besedna neizraženost v skladenjskem vzorcu sicer predvidenega; npr. <Sedele smo za mizo.> izpuščeno <me>; sopom. elizija
jakostni naglas	naglas, ki naglašene zloge loči od nenaglašanih z večjo glasnostjo (z močnejšim izdihom); v nasprotju s tonemskim naglasom
jakostna členitev	členitev na nenaglašene in naglašene besede in zloge, oslabiljeno in navadno naglašene besede, poudarjene in nepoudarjene naglašene in nenaglašene zloge ali dele stavka oz. povedi; v besedi je jakostno najizrazitejši del jakostno težišče.
jakostna stopnja	v nevtralni jakostni členitvi se ločijo: 1. nenaglašenost, 2. oslabiljena naglašenost, 3. navadna naglašenost, 4. poudarjena naglašenost; nenevtralna jakostna stopnja nastopi pri vpitju ali šepatenju
jakostni naglas	naglasna izrazitost, dosežena z večjo jakostjo pri izgovoru naglašene zloga v primerjavi z nenaglašenim zlogom; je značilen za večino slovenskih narečnih skupin; sopom. dinamični naglas
jakostno težišče	najizrazitejši, tj. naglasno neoslabiljeni zlog besede ali pa poudarjeni zlog člena govorne verige; predstavlja tudi intonacijsko težišče
kadenca	zadnji del pripovedne končne stavčne intonacije, usmerjen proti spodnji tonski legi, ko gremo z glasom navzdol
kakovost samoglasnika	barvna lastnost samoglasnika, odvisna od mesta največjega pridviga jezične ploskve proti raznim točkam trdega ali mehkega neba; npr. pri ozkem ali širokem [E], pri [i] in [u]
klitika	gl. naslonka

knjižni govor	govorno uresničevanje besedil v knjižnem, zlasti zbornem jeziku
knjižni pogovorni jezik	manj zahtevna oblika knjižnega jezika, uporabljana zlasti v neformalnem sporočanju: v pogovoru ali v govornih nastopih pred manjšim poslušalstvom
končaj	končni del besede, ne glede na morfemskost; besede <kost>, <most>, <ost> imajo isti končaj <-ost>, besedi <krepi> in <oči> pa končaj <i> 2. končaj
konzontan	gl. soglasnik
krajšanje	prozodični potek, pri katerem se glas (samoglasnik, bolj poredko soglasnik) skrajša
kratica	samostalni, nastal iz začetnih delov večbesednega poimenovanja; prim. <TOMOS> iz <Tovarna Motorjev Sežana>
krativec	naglasno znamenje, sopom. gravis
ločilo	pisavno znamenje (nečrkovno, neštevnico, nesimbolno), zaznamujoče skladijsko razmerje, uporablja pa se tudi za neskladijsko.
medsebojno vplivanje glasov	glasovi se v govoru med seboj vedno prilagajajo glede na sosednje glasove; posebno zanimivo je tisto vplivanje, katerega posledica je fonemska različica ali alofon ali celo premena enega fonema z drugim. Glavni vrsti vplivanja sta: prilikovanje ali asimilacija ter razlikovanje ali disimilacija.
mehčani soglasnik	palatizirani soglasnik
mehkonebnojezičen	soglasnik izgovarjen z jezično ploskvijo proti mehkeemu nebu, npr. [k], [g], [h]
mesto naglasa	zlog, na katerem im naglašena beseda jakostno ali tonsko izrazitost, ki vpliva na vse ostale zloge. Ločimo mesti naglasa na zadnjem zlogu (oksitoneza), ne na zadnjem zlogu (baritoneza), na predzadnjem zlogu (paroksitoneza), na predpredzadnjem zlogu (proparoksitoneza). Naglasno mesto je lahko obvezno (stalni naglas), npr. v francoščini vedno na zadnjem zlogu, v poljščini na predzadnjem, v makedonščini na predpredzadnjem zlogu, v češčini na prvem zlogu, ali pa neobvezno (prosti naglas), npr. v slovenščini, ruščini. Naglasno mesto je lahko tudi delno omejeno (v hrvaščini ga ni na zadnjem zlogu)
morfem	najmanjša skupina fonemov ali alofonov, ki ima svoj pomen, npr. lip - in -a v besedi lipa

morfonologija	veda, ki proučuje glasovne in naglasne značilnosti morfema oziroma morfemov besede
naddolžina	kar se pri govoru zategne, npr. izglasni samoglasnik na koncu govornega člena, zlasti pred premorom na koncu sporočila
naglas	jakostna ali tonska izrazitost enega zloga v besedi nasproti drugemu; slovenski knjižni jezik dopušča jakostni in tonemski naglas; sopom. akcent
naglasno mesto	gl. mesto naglasa
naslonka	beseda ali njena oblika, ki ni naglašena, npr. predlog, veznik, členek, pomožne glagolske oblike <sem>, <bi>, <bom>, zaimenske oblike <me>, <mi>, <jo>; sopom. klitika, breznaglasnica
nenaglasnica	beseda, ki je sicer naglašena, vendar v dani zvezi nepoudarjena, npr. beseda <dobro> v pozdravu <Dobro jutro.>
nepisani polglasnik	govorjeni polglasnik pred [r], ki ni ob samoglasniku istega zloga; npr. <smrt>, <rdeč>, <zardeti>
neskladenjska raba ločil	raba ločil za zaznamovanje oblikoslovnih ali oblikoglasnih lastnosti; npr. pika zaznamuje za številko vrstilstnost ali kratnost ipd.
nezvočniki	glasovi najmanjše odprtostne stopnje; pred seboj zahtevajo po zvonečnosti enak soglasnik [p], [b], [f], [t], [d], [s], [z], [ts], [S], [Z], [tS], [dZ], [k], [g] in [h]
netrajni soglasnik	soglasnik, ki se ne da vleči; v slovenskem knjižnem jeziku so to [p], [t], [k], [b], [d], [g], [ts], [tS], [dZ]; sopom. netrajnik
nosilec zloga	v dani okolici glas največje zvonečnosti, večinoma je to samoglasnik, lahko pa tudi zvočnik ali nezvočnik; pri dvoglasnikih sta to oba dela
oblikoglasen	nanašajoč se na glasovno naglasno podobo morfema ali besede; sopom. morfonološki
obvezno naglasno mesto	slovenske besede se, če so večzložne, praviloma naglašujejo le na enem zlogu; o mestu naglasa nas uči teorija o predvidljivosti naglasnega mesta
odprti zlog	zlog, katerega zvonečnost je proti koncu največja, ker se končuje s samoglasnikom
odprtostna stopnja	razdalja med zgornjim in spodnjim delom govorne cevi; največjo odprtostno stopnjo imajo samoglasniki, srednjo zvočniki, najmanjšo nezvočniki
oksiton	jezikoslovni izraz za besedo, ki ima naglas na zadnjem zlogu (enozložna ali večzložna)

osnova	del pregibne besede pred končnico ali del glagola, vključno z glagolsko pripono, npr. <kupova-> v <kupovati>
osnovna frekvenca	frekvenca tresenja glasilk pri tvorjenju zvonečih glasov
osnovna perioda	čas, ki preteče med dvema zaporednima tresljajema glasilk pri tvorjenju zvonečih glasov
ostrivec	naglasno znamenje v obliki krajše črtice nad črko od leve spodaj desno navzgor, npr. v besedi [pét]
padajoča intonacija	tonski potek glasovnega gradiva od začetne višine proti točki spodnjega roba srednje lege glasu; izrazito navzdol se obrne po težiščnem zlogu, zlasti še, če je ta poudarjen (pri tonemskem naglasu to velja za cirkumflektirano težišče).
padajoča polkadenca	polkadenca, kakor je padajoča antikadenca, le da proti koncu nekončnega intonacijskega člena, pisno zaznamovanega s podpičjem ali dvopičjem
paroksiton	beseda, naglašena na predzadnjem zlogu
podstava	del besede ali oblike pred obrazilom; prim. <kup-> v besedi <kupujem>
polkadenca	glava tonskega poteka nekončnega člena stavčne intonacije
ponaglasen	ki se nahaja za naglašnim zlogom; sopom. posttoničen
poudarek	stavčnofonetična izrazitost kakega dela stavka ali stavčne zveze, npr. v stavku <Naša Maja gre jutri na izlet.> želimo poudariti, da gre za Majo in ne za Matejo ali Mojco; na poudarjenem zlogu je težišče stavčne intonacije.
poudarjeno težišče	težišče intonacije oz. intonacijskega segmenta: zlog, ki pridobi izjemno izrazitost zaradi povečane jakosti izgovora, npr. <KAJ sem ti storila?>; v nasprotju z navadnim težiščem, odvisnim od položaja zloga v intonacijskem segmentu; protipom. nepoudarjeno težišče
poved	del besedila, ki se v pisavi na začetku zaznamuje z veliko začetnico, na koncu s piko, vprašajem ali klicajem, v govoru pa z razmeroma visokim začetkom intonacije ter s kadenco ali antikadenco na koncu; npr. <Dobro jutro!> ali <Lepo je, da ste prišli!>
povedna intonacija	stavčna intonacija pripovedne povedi: ima kadenčno intonacijsko glavo, pod pogojem, da poved ni bila izražena čustveno; v tem primeru se povedna intonacija zamenja z vzklično
prednaglasen	ki se nahaja pred naglašnim zlogom
predpona	morfem levo od korena
predslonka	nenaglašena slovnična beseda ali oblika, ki tvori izgo-

	vorno celoto z naslednjo naglašeno besedo, npr. predlog, veznik, členek; sopom. proklitika
premor	stavčnofonetični pojav, da se govorna veriga na določenih mestih pretrga. Ta mesta so navadno na koncu povedi: med prirednimi deli povedi, med pristavkom, izpostavkom, pastavkom ipd.; sopom. pavza
prilikovanje	premena enega glasu z drugim pod vplivom glasovnega konteksta; sopom. asimilacija
pripona priponsko obrazilo	morfem desno od korena; prim. miz -ic -a , kup -ov -a -ti morfem, ki se pri tvorbi nove besede dodaja desno od podstave, lahko ima tudi končnico; prim. <ic(a)> v besedi <mizica>
proklitika	gl. predslonka
proparoksiton	beseda, naglašena na predpredzadnjem zlogu
prozodem	enota prozodije, npr. naglas (jakostni, tonemski), v okviru besede, stavka ali povedi
prozodija	nadsegmentne lastnosti govorne besede od glasu preko besede do povedi, npr. naglasnost, mesto naglasa, vrsta naglasa, intonacija, stavčna fonetika
ravna intonacija	zaporedni takti se od leve na desno ne spuščajo
rep intonacije	potežiščni del intonacijske glave
ritem	menjavanje naglašanih in nenaglašanih zlogov, posledica stavčnofonetičnih značilnosti
samoglasniki	glasovi z največjo odprtostno stopnjo
samoglasniška redukcija	gl. upad samoglasnika
SAMPA	SAM Phonetic Alphabet, računalniško berljiva mednarodna fonetična abeceda; definirana v okviru ESPRIT projekta 1541 – SAM (Speech Assessment Methods)
SAMPROSA	SAM Prosodic Transcription, računalniško berljiva zbirka znakov za opis prozodičnih parametrov govora
soglasniki	glasovi z manjšo odprtostno stopnjo kot samoglasniki; v navadnem besedju niso zlogotvorni; po odprtostni stopnji se delijo na zvočnike in na nezvočnike
soglasniški sklop	zveza več soglasnikov v isti besedi ali na sklopu besed; prim. strah, od hiše, prav k tebi
stavčna fonetika	veja glasoslovja, ki proučuje slušne lastnosti skladenjskih enot, tj. besednih zvez, stavkov, povedi; te lastnosti so poudarjanje, členitev s premori, intonacija, register, hitrost in glasovno barvanje
stavčna intonacija	podoba tonskega poteka v povedi ali njenem delu; končna se končuje s kadenco ali antikadenco, nekončna pa s polkadenco

stranski naglas	drugi naglas v besedi
števka	znamenje za števila od nič do devet; npr. 15 sestoji iz dveh števk, iz 1 in 5
šumnik	nezvočnik, izgovarjan s hrbtom jezika, približanim trdemu nebu in v paru s sičnikom; v slovenskem knjižnem jeziku [S], [Z], [tS] in [dz]
takt	v okviru tega dela je takt definiran kot skupek besede in brezglasnic, ki se na besedo naglasno naslanjajo
telo intonacije	lik, ki ga v okviru intonacijskega sklopa tvorijo naglašeni zlogi tega člena, vključno s težiščem glave intonacije
tempo artikulacije	tempo artikulacije je definiran kot razmerje med številom izgovarjanih zlogov in časom artikulacije, kjer so izključeni vsi premori, ne glede na njihov njihov značaj ali trajanje
tempo govora	gl. hitrost govora
težišče intonacije	naglašeni ali poudarjeni zlog, s katerim se začenja glava intonacije; težišču sledi rep intonacije; v težišču se intonacija obrne: pri povedni intonaciji težišče tonsko drugačno kot pri vprašalni, pri vprašalni je lahko visoko ali nizko, pri poudarjanju pa navzgor ali navzdol premaknjeno.
tonem	pomenskorazločevalni ton v besedi (v slovenščini le naglašeni); prim. cirkumfleks in akut
tonemski naglas	naglas, pri katerem se ustrezna izrazitost dosega s tonom; v nasprotju z jakostnim naglasom; tonsko naglašeni samoglasniki so lahko izgovarjani višje (cirkumfleksirani) ali nižje (akutirani)
trajnik	soglasnik (zvočnik ali nezvočnik), ki se da pri izgovarjavi vleči; v slovenskem knjižnem jeziku so to zvočniki <m n v j r l> in nezvočniki <f s z š ž h>; nasprotje netrajnik
trup intonacije	lik, ki ga v okviru intonacijskega sklopa tvori naglašeni zlogi tega sklopa, vključno s težiščem glave intonacije; če se glava začenja s prvim naglašenim oz. poudarjenim zlogom segmenta, je trup ničti. Pomenskorazločevalne vloge trup intonacije nima, je pa različen pri različnih vrstah intonacije: povedne, vprašalne, vzklične; sopom. telo intonacije
upad samoglasnika	izguba prvotne izrazitosti samoglasnika; v skrajnjem primeru izgine tudi iz besede; prim. [tu:di] gre v [tu:d@] li celo v [tud]; sopom. samoglasniška redukcija
ustničnik	soglasnik, pri katerega izgovoru so udeležene ustnice

ustničnoustnični soglasnik	soglasnik, izgovarjan s primikom spodnje ustnice k zgornji; v slovenskem knjižnem jeziku /m/ in nepredsamoglasniške variante /v/
večnaglasnost	pri besedah z več naglasi, v slovenščini so to prave sestavljenke, prim. <netalentiran> in zloženke določenih vrst <bledosiv>
višina osnovnega tona	višina, na kateri poslušalec zaznava govor, je povezana z osnovno frekvenco govora
vokal	gl. samoglasnik
vprašalna intonacija	intonacija, nasprotna povedni; v težišču se ne obrne navzdol proti tonskemu dnu registra, ampak navzgor, ali pa se celo težišče prenese navzgor
vzklična intonacija	pod vplivom čustvenosti spremenjena povedna ali vprašalna intonacija, npr. z večjim ali izravnanim nizkim intervalnim obsegom na višjem ali nižjem registru, pogosto s povečano jakostjo
zamolk	figura, za katero je značilno, da se začeti skladenjski vzorec ne izpelje do konca; prim. <Tiho bodi, če ne...>; so-pom. aposiopeza
zapisovanje glasov	govorjeno besedilo lahko zapisujemo fonetično, tj. z glasovnimi različicami ali alofoni med oglatimi oklepaji ali fonološko (s fonemi) med poševnimi oklepaji
zapisovanje intonacije	poenostavljeno intonacijo podajamo s prerisovanjem tonskega poteka pri samoglasnikih, pri čemer posebej zaznamujemo naglašene od nenaglašanih. Še bolj poenostavljeno intonacijo prikazujemo s črtico za naglašeni in s piko za nenaglašeni zlog, pri tem pa seveda upoštevamo relativne višinske razlike enih in drugih.
zborni govor	višja, zahtevnejša oblika knjižnega jezika, zlasti govorenega
zborni jezik	stroga, širša, varianta knjižnega jezika s strogo določenimi izgovornimi in oblikoslovnimi posebnostmi
zborni knjižni jezik	zborna različica knjižnega jezika v nasprotju s splošnim pogovornim jezikom ali knjižnim pogovornim jezikom nezvočnik, pri katerem je govorna cev najprej zaprta, potem pa se počasi odpira; v slovenskem knjižnem jeziku so to [ts], [tʂ] in [dʒ]
zlitnik	izgovarjan s spodnjo ustnico ob zgornjih sekalcih; v slovenskem knjižnem jeziku [f] in [v]
zobnoustničen	izgovarjan s spodnjo ustnico ob zgornjih sekalcih; v slovenskem knjižnem jeziku [f] in [v]
zvenečnost	lastnost glasu, dosežena s tresenjem glasilk; v slovenskem knjižnem jeziku so zveneči vsi samoglasniki in

zvočnik

zvočniki (od teh ima samo /v/ nezvenečo varianto: [W])
ter nezvočniki [b], [d], [g], [dZ], [z] in [Z]
soglasnik srednje odprtostne stopnje, srednje zvonkosti,
trajen, brez pomenskorazločevalnih parov glede na zve-
nečnost; v slovenskem knjižnem jeziku so zvočniki: no-
snika [m] in [n], drsnika [v] in [j] in jezičnika [r] in [l]

Sinteza govora je proces umetnega ustvarjanja govora, ki ga lahko uporabljamo za najrazličnejše namene [Dutoit 93a, str. 19–20]. Sistemi za pretvorbo besedila v govor omogočajo posredovanje informacij, kot so teletext, elektronska pošta, faksi in druga besedila, brez potrebe po njihovi vidni predstavitvi, kar je še zlasti uporabno za slepe in slabovidne osebe. Nadalje se uporabljajo v sistemih za avtomatsko posredovanje informacij, tj. v tako imenovanih interaktivnih govornih odzivnikih [Sorin 94], kjer bi bilo tovrstno delo za človeka enolično in utrujajoče.

Sistem za sintezo slovenskega govora sestavlja več medsebojno neodvisnih modulov, ki so povezani v hierarhično arhitekturo. Vhod v sintetizator govora predstavlja poljubno besedilo, shranjeno v računalniški obliki v zapisu ASCII. Besedilo se sprva pretvori v ustrezeni fonetični zapis. Sledi določanje osnovne frekvence in trajanja za posamezne glasove, nato se oblikuje še celotna stavčna intonacija. Sam postopek sinteze govornega signala temelji na lepljenju osnovnih govornih enot s pomočjo algoritma TD-PSOLA [Moulines 90]. Za osnovne govorne enote smo izbrali difone. Predvajanje digitalnega zapisa govornega signala poteka preko zvočne kartice računalnika.

Vhod v sintetizator govora predstavlja poljubno besedilo v slovenskem jeziku, zapisano v formatu ASCII. Pretvorba besedila v fonetični prepis poteka v dveh korakih: predobdelava besedila in pretvorba grafemskega zapisa besedila v fonetični prepis.

Sprva izvedemo predobdelavo vhodnega besedila. V tem koraku se iz besedila odstranijo vsi odvečni simboli, ki ne vplivajo na izgovarjavo tega besedila, prav tako se zaporedja števk razvijejo v ustrezne števnike, glavne ali vrstilne ... Okrajšave in kratice se zapišejo v polni obliki. Za vsako ločilo se ugotovi, ali nastopa v skladenjski ali v neskladenjski rabi.

Sledi določanje izgovarjave besede. Če besede ne najdemo v slovarju pogostih izgovarjav (trenutno obsega okoli 40.000 najbolj pogostih slovenskih besed), ji moramo izgovarjavo določiti samodejno. Najprej napovemo naglasno mesto, nato besedo preslikamo v njen fonetični prepis.

Za slovenski jezik je značilno prosto mesto naglasa, saj se ta lahko pojavi na prvem, zadnjem, predzadnjem ali predpredzadnjem zlogu. Mesto naglasa je določeno za vsako besedo posebej in se ga naučimo hkrati z učenjem jezika in besed. Vendarle obstajajo pravila, ki delno določajo mesto naglasa [Toporišič 91]. Ta pravila smo uporabili za samodejno določanje mesta naglasa v besedi, vendar ne zaobjamejo vseh besed. Zato smo slovar pogostih izgovarjav statistično analizirali in poiskali pravila, ki napovedujejo naglasno mesto v besedi glede na število zlogov v besedi. Z dobljenimi pravili smo določili najbolj verjetno naglasno mesto v besedi.

Za pretvorbo besed v ustrezna zaporedja fonetičnih simbolov smo uporabili preko 140 produkcijskih pravil za posamezne glasove oziroma skupine glasov. Pravila smo po-

novno optimizirali na slovarju pogostih izgovarjav. Pri določanju izgovarjave besed smo upoštevali tudi koartikulacijske vplive na stikih med besedami.

Pravilnost izbire prozodičnih parametrov močno vpliva na kakovost sintetiziranega govora. Postopek nastavljanja prozodičnih parametrov se deli na dva poglavita koraka: modeliranje trajanja in modeliranje osnovne frekvence.

Za nastavljanje trajanja govornih enot za slovenski jezik uporabimo *dvostopenjski pristop* kontrole trajanja [Epitropakis 93c]. Sprva govornim enotam v trajanju enega takta priredimo inherentne ali *intrinzične dolžine*, ki jih dobimo kot vsoto intrinzičnih dolžin glasov, vsebovanih v taktu. Ko se takti vključujejo v večje govorne enote (fraze, stavke), se v skladu z zahtevami višjenivojskih prozodičnih pojavov *skrajšujejo* ali *podaljšujejo* in privzamejo *ekstrinzično dolžino*. Pri določanju ekstrinzične dolžine takta upoštevamo različne skladijske parametre, kot so položaj takta v frazi, izbrana hitrost govora in dolžina takta, izražena s številom zlogov v taktu. Sledi postopek prilagajanja intrinzične dolžine takta na ekstrinzično dolžino. To dosežemo s podaljševanjem oz. skrajševanjem intrinzične dolžine takta. Pri tem je treba upoštevati, da se vsi glasovi znotraj takta ne skrajšujejo oz. podaljšujejo v enaki meri. Iz predvidenega trajanja oz. hitrosti zloga, besede ali takta izpeljemo trajanje glasu ali pa trajanje glasu popravimo glede na znani tempo artikulacije izgovorne celote.

Za preklap med obema nivojema za razliko od [Ferreira 93], ki za preklopno govorno enoto jemlje besedo, izberemo *takt*. S tem se izognemo nevarnosti, da bi breznaglasnice, ki so pogosto kratke, enozložne besede, obravnavali kot samostojne enozložne besede, katerim pripada razmeroma nizek tempo artikulacije.

Modela za določanje intrinzičnih dolžin glasov in ekstrinzičnih dolžin taktov temeljita na rezultatih, ki smo jih dobili z merjenjem trajanja glasov v izbranih ločeno izgovarjanih besedah in tudi na tekočem govoru za prvi model, in na rezultatih meritev trajanja taktov na tekočem branem govoru za drugi model. Postopek prilagajanja skupnega trajanja glasov v taktu zahtevani dolžini takta temelji na rezultatih, ki smo jih dobili z meritvami spreminjanja trajanja posameznih glasov ob različnih hitrostih govora.

Za pravilno upoštevanje rezultatov modeliranja trajanja različnih govornih enot iz obeh modelov smo sestavili postopek, ki intrinzično dolžino takta, določeno z vsoto intrinzičnih dolžin posameznih glasov v taktu, prilagodi zahtevani ekstrinzični dolžini, in pri tem vzpostavi pravilna razmerja med trajanji posameznih glasov znotraj takta. Pri tem je treba v postopku pravilno upoštevati različno relativno spreminjanje trajanja posameznih glasov ob spremembi trajanja celotnega takta [Gros 97e].

V slovenskem knjižnem jeziku naglašujemo na dva načina, jakostno in tonemsko [Toporišič 91, str. 55]. Pri tonemskem naglaševanju je za naglas značilno naraščanje tona znotraj naglašene zloga. Po dosegu tonskega vrha pa sledi upadanje, ki je odvisno od vrste besede (oksiton ali bariton) ter od vrste naglasa (akut ali cirkumfleks).

Zaradi narave oblikovanja poteka osnovne frekvence v slovenskem jeziku smo pri nastavljanju osnovne frekvence sintetiziranega govora uvedli *dvostopenjski princip* in razdelili oblikovanje poteka osnovne frekvence na dva dela: oblikovanje besednega naglasa in oblikovanje stavčne intonacije.

Besedni naglas oblikujemo tako, da vsaki naglašeni besedi oz. taktu predpišemo enega izmed obeh značilnih tonemskih naglasov in nato oblikujemo potek osnovne frekvence za to besedo oz. takt. Oblikovanje besednega naglasa poteka v petih korakih: nastavljanje začetnih vrednosti osnovnih frekvenc, nastavljanje skoka pri tonemskem naglasu, omejitev skokov, interpolacija vmesnih vrednosti osnovne frekvence in razgibanje osnovnih frekvenc.

Pri nastavljanju začetnih vrednosti samoglasnikov smo, v skladu z ugotovitvami Srebotove [Srebot 88, str. 33–39], upoštevali naslednje dejavnike: inherentno višino samoglasnika, položaj zloga, v katerem se samoglasnik nahaja (začetni, končni, sredinski), odprtost zloga (odprt ali zaprt) in naglašenost zloga (naglašen, ponaglasni, prednaglasni).

V slovenščini sta najbolj pogosta dva tonemska naglasa, to sta akut in cirkumfleks. Sprva taktu priredimo vrsto naglasa. Nato za vsak takt oblikujemo potek osnovne frekvence po enem izmed *značilnih vzorcev osnovne frekvence*. Pri tem upoštevamo vrsto naglasa (akut ali cirkumfleks), število zlogov v taktu in položaj naglašene zloga v besedi oz. taktu (oksiton ali bariton).

Nastavljanju tonemskega naglasa besed smo posvetili večjo pozornost kot oblikovanju stavčne intonacije. Superpozicijo besednega naglasa in stavčne intonacije dosežemo tako, da vrednosti osnovne frekvence skaliramo s faktorjem, ki podaja vpliv stavčne intonacije na osnovno frekvenco v besedi. S tem hkrati povečamo tonski obseg sintetičnega govora. Vrsto intonacijskega vzorca, ki ga priredimo posamezni frazi, določamo na podlagi ločila na koncu fraze.

Za nastavitev stavčne intonacije smo uporabili preprost postopek, ki ne zahteva predhodne skladijske oziroma pomenske razgradnje stavka [Sorin 87]. Pri določanju nastavitvenih pravil smo se naslonili na predhodne ugotovitve jezikoslovcev o stavčni intonaciji za slovenski govor [Toporišič 69].

Za samo sintezo govornega signala smo uporabili postopek, ki temelji na združevanju osnovnih govornih enot in hkrati omogoča spreminjanje prozodičnih lastnosti glasov. Na vhodu modula se preberejo podatki o zaporedju glasov, ki ga želimo sintetizirati, skupaj z želenimi vrednostmi osnovne frekvence in časa trajanja za posamezen glas. Na izhodu modula dobimo sintetiziran govor.

Za osnovne govorne enote smo izbrali difone. Difon predstavlja sklop dveh sosednjih glasov oz. alofonov, in sicer drugo polovico prvega glasu in prvo polovico drugega glasu. Na ta način difon vsebuje informacijo o glasovnem prehodu enega glasu v drugega. V bazi difonov, ki smo jo posneli, se nahaja 1024 difonov. Posneti difoni so bili izrezani iz sredine brezpomenskih besed, logatomov, izgovarjanih z razmeroma enakomerno intonacijo. Besede smo izbrali tako, da se želeni difon nahaja v fonetično čim bolj nevtralnem okolju z ostalimi glasovi.

Za združevanje osnovnih govornih enot smo uporabili postopek TD-PSOLA [Moulines 90]. Postopek TD-PSOLA je neparametrična metoda in zato ne zahteva poznavanja modela produkcije govornega signala ter v celoti ohrani kakovost shranjenih delov

signala. Dovoljuje kakovostne spremembe osnovne frekvence in časov trajanja osnovnih govornih enot neposredno na časovnem signalu.

Da bi ugotovili kakovost, naravnost in možnost uporabe umetno sintetiziranega govora, smo v Laboratoriju za umetno zaznavanje organizirali preizkus sintetizatorja, pri katerem je sodelovalo 21 poslušalcev, starih med 19 in 45 let. Del preizkusa smo zasnovali po priporočilih mednarodne zveze za telekomunikacije ITU, v preostalem delu preizkusa pa smo želeli dobiti oceno posameznih nastavitvev sintetizatorja govora.

Prvi del preizkusa je pokazal, da je sintetični govor poslušalcem razumljiv. Poslušalce smo povprašali, ali se jim zdi, da bi bil sintetizator govora v tej obliki primeren za samodejno podajanje informacij po telefonu. Več kot dve tretjini poslušalcev sta menili, da bi bil primeren, manj kot ena tretjina je menila, da ne bi bil primeren. Kar nekaj poslušalcev, ki so obkrožili 'ne', je po preizkusu neuradno pripomnilo, da so tako oceno dali predvsem zato, ker se jim zdi neprimerno, da bi stroji ljudem odvzemali delovna mesta, in ne zato, ker bi se jim zdela kakovost sintetiziranega govora prenizka.

V drugem delu testa so poslušalci kakovost sintetičnega govora uvrstili med naravni govor, pošumljen z moduliranim Gaussovimi šumom v razmerju signal : šum 5 dB in 10 dB. V tretjem delu testa so poslušalci ocenjevali različne nastavitve parametrov sintetizatorja govora.

Kakovost končnega sintetiziranega govora je odvisna od vseh modulov v sistemu. Preizkus naravnosti in razumljivosti je pokazal, da je umetno sintetiziran govor primeren za poslušanje. Najšibkejšo točko sistema trenutno predstavljajo napake zaradi nepravilne določitve naglasnega mesta v besedi, kar prispeva k povečani nerazumljivosti in nenaravnosti govora.

AUTOMATIC TEXT-TO-SPEECH SYNTHESIS (SUMMARY)

The present volume describes efforts oriented towards automatic production of intelligible Slovene speech from arbitrary Slovene texts. It is based on a PhD thesis [Gros 97e] that has been updated and rewritten in order to be made comprehensible to a more linguistically oriented reader.

We continue with a description of the contents of the volume.

Text-to-speech synthesis (TTS) enables automatic conversion of any available textual information into its spoken form. In the Laboratory of Artificial Perception at the University of Ljubljana, we started on text-to-speech synthesis in 1995. In the following year we presented the first PC-based TTS system for the Slovene language [Gros 95]. We used it as a reference system for further improvements. In the recent version of our TTS system S5 we implemented a novel procedure for the determination of prosodic parameters, added a pronunciation dictionary and improved text normalization [Gros 97e].

The S5 TTS system has already been implemented in different applications: S5 as a stand-alone TTS system (reading machine), S5 providing the spoken output in the SQEL speech recognition and dialog system for automatic airline timetable retrieval, S5 integrated into a special application – HOMER, a reading system for the blind and partially sighted people, combined with optical character recognition devices and voice control and a TTS web server at <http://glas.fe.uni-lj.si>.

The input text is transformed into its spoken equivalent by a series of modules, which we describe in detail. A grapheme-to-phoneme or grapheme-to-allophone module produces strings of phonetic symbols based on information in the written text. The problems it addresses are thus typically language-dependent. A prosodic generator assigns pitch and duration values to individual phones. Final speech synthesis is based on diphone concatenation using TD-PSOLA [Moulines 89].

Input to the S5 system is free text. For the time being, input text should be stored in ASCII format; currently we are expanding the input possibilities so that in future it may come from other programs or marked regions on the computer screen.

Input text is translated into a series of allophones in two consecutive steps. First, input text normalization is performed. Abbreviations are expanded to form equivalent full words using a special list of lexical entries. The text normalizer converts further special formats, like numbers or dates, into standard grapheme strings. The rest of the text is segmented into individual words and basic punctuation marks.

Next, word pronunciation is derived, based on a user-extensible pronunciation dictionary and letter-to-sound rules. The dictionary covers over 40,000 of the most frequent inflected word forms.

In case the dictionary derivation fails, words are transcribed using automatic lexical

stress assignment and letter-to-sound rules. However, as lexical stress in Slovene can be located almost arbitrarily on any syllable, this step can introduce errors into the pronunciation of words. For words which do not belong to these categories, the most probable stressed syllable is predicted using the results obtained by a statistical analysis of stress position depending on the number of syllables within a word.

Finally, a set of over 140 context-dependent letter-to-sound rules translate each word into a series of allophones.

A number of studies suggest that prosody has a great impact on the intelligibility and naturalness of speech perception. Only the proper choice of prosodic parameters, given by sound duration and intonation contours, enables the production of natural-sounding high quality synthetic speech.

Prosody generation in S5 consists of four phases: intrinsic duration assignment, extrinsic duration assignment, modelling of the intra word F_0 contour and assignment of a global intonation contour.

A speech database consisting of isolated words, carefully chosen by phoneticians, was recorded in order to study different effects on phone duration and fundamental frequency, which operate on a segmental basis. Vowel duration and F_0 were studied in different types of syllables: stressed/unstressed, open/closed. Consonant duration was measured in CC and VCV clusters.

Another large continuous speech database was recorded to study the impact of speaking rate on syllable duration and duration of phones. A male speaker was instructed to pronounce the same material at different speaking rates: at a normal, fast and slow rate. Thus context, stress and all other factors were kept identical to every realisation of the sentence. As a result, pair-wise comparisons of phone duration could be made.

The effect of speaking rate on phone duration was studied in a number of ways. An extensive statistical analysis of lengthening and shortening of individual phones, phone groups and phone components, like closures or bursts was performed, the first of the kind for the Slovene language.

Articulation rate expressed as the number of syllables or phones per second, excluding silences and filled pauses, was studied for the different speaking rates. In other studies, articulation rate is usually determined for speech units with the length of individual words or entire phrases. We studied the articulation rate of words along with their associated cliticised words at different positions within a phrase: isolated, phrase initial, phrase final and nested within the phrase. The articulation rate increases with longer words, as average syllable duration tends to decrease with more syllables in a word. The articulation rate immediately after pauses is higher than the one prior to pauses.

A set of measurements was made in order to define four typical intonation contours based on four Slovene basic intonation types. Read newspaper articles were processed by an AMDF pitch extractor. Then, a manual piecewise linearization of F_0 curves into pitch contours was performed. Our interest was to detect typical prosodic segments by means of F_0 contours.

Similarly to [Epitropakis 93b], our two-level duration model first determines the words'

intrinsic duration, taking into account factors relating to the phone segmental duration, such as: segmental identity, phone context, syllabic stress and syllable type: open or closed syllable.

Furthermore, the extrinsic duration of a word is predicted, according to higher-level rhythmic and structural constraints of a phrase, operating on a syllable level and above. Here the following factors are considered: the chosen speaking rate, the number of syllables within a word and the word's position within a phrase, which can be isolated, phrase initial, phrase final or nested within the phrase.

Finally, intrinsic segment duration is modified, so that the entire word acquires its pre-determined extrinsic duration. It should be noted that stretching and squeezing does not apply to all segments equally. Stop consonants, for example, are much less subject to temporal modification than other types of segments, such as vowels or fricatives.

Therefore, a method for segment duration prediction was developed, which adapts a word with an intrinsic duration t_i to the determined extrinsic duration t_e , taking into account how stretching and squeezing apply to the duration of individual segments [Gros 97e].

The reliability of our two-level prediction method was evaluated on a speech corpus consisting of over 150 sentences. The predicted durations were compared to those in the same position in natural speech. Natural duration variation was evaluated by averaging the duration differences for words, which occurred in the corpus several times, in the same phonetic environment and in the same type of phrase. Standard deviation of the difference between natural and predicted duration difference is 15.4 ms for a normal speaking rate, and even less for stressed phonemes, the duration of which is of crucial importance to the perception of naturalness of synthetic speech.

Since the Slovenian language has been defined as a pitch accent language, special attention was paid to the prediction of tonemic accents for individual words. First initial vowel fundamental frequencies were determined according to previous measurements as suggested by [Srebot 88], creating the F_0 backbone. Each stressed word was assigned one of the two tonemic accents, characteristic for the Slovene language. The acute accent is mostly realized by a rise on the post tonic syllable, while with the circumflex the tonal peak usually occurs within the tonic. Five typical F_0 patterns were chosen from the variety of F_0 patterns described. Finally a linear interpolation between the defined F_0 values was performed.

We used a relatively simple approach for prosody parsing and the automatic prediction of Slovenian intonational prosody which makes no use of syntactic or semantic processing, but rather uses punctuation marks and searches for grammatical words, mainly conjunctions which introduce pauses. We considered it more important to predict the word F_0 contour modeling the tonemic accent as reliably as possible than to explore sentence intonation.

Once appropriate phonetic symbols and prosody markers are determined, the final step within S5 is to produce audible speech by assembling elemental speech units. This is

achieved by taking into account computed pitch and duration contours, and synthesizing a speech waveform.

A concatenative synthesis technique was used. The TD-PSOLA scheme enables pitch and duration transformations directly on the waveform, at least for moderate ranges of prosodic modifications [Moulines 90] without considerably affecting the quality of the synthesized speech.

Diphones were chosen for concatenative speech units as a compromise between the size of the unit inventory, the complexity of the concatenation rules and the resulting speech quality.

The adequacy of the spoken output was evaluated by several subjective tests as recommended by the International Telecommunication Union (ITU). The adequacy of the S5 system was evaluated in terms of acceptability and intelligibility. The experiment was performed in laboratory conditions with 21 subjects within an age span of 19 and 45 years, 10 of them being female. It was conceived according to various ITU-T recommendations, describing methods for subjective performance assessment of the quality of speech voice output devices.

The test was divided into three parts. The first part was to evaluate whether the quality of the synthetic speech was sufficiently high for a real application of the system in an automatic information retrieval system. The subjects were asked to fill in different templates related to the chosen application domain based on the information they heard.

The second part of the test served to compare several features describing the synthetic voice quality to those describing the quality of natural speech distorted with different levels of gaussian noise. The synthetic speech received a mean opinion score, which was between distorted natural speech with a SNR ratio of 5dB and 10dB.

In the third part of the test, different methods for prosody assignment were evaluated. The majority of the subjects estimated the synthetic speech produced by S5 to be pleasant and quite natural sounding, sufficiently rapid and not over-articulated.

S5 is a text-to-speech system for the Slovene language, capable of synthesizing intelligible continuous speech from an arbitrary Slovene input text. Further improvement of intelligibility and naturalness depend in particular on proper lexical stress assignment and a more sophisticated generation of F_0 prosodic parameters. The first attempts at developing a diphone-based synthesis system for the Slovene language are promising, so that further work on improving individual parts of the system is encouraged.

Predgovor	5
1. Uvod	7
1.1 Psihološke in fiziološke osnove govora	8
1.2 Postopki sinteze govora	9
1.3 Sinteza govora za slovenski jezik	9
1.4 Zasnova sistema za sintezo slovenskega govora	10
2. Predobdelava vhodnega besedila	12
2.1 Črkovni grafemi	13
2.2 Ločila	14
2.3 Števke	15
3. Pretvarjanje grafemskega zapisa besed v fonetičnega	19
3.1 Določanje izgovarjave besed	20
3.2 Iskanje besed in kolokacij v slovarju najbolj pogostih izgovarjav	21
3.3 Pretvarjanje grafemskega zapisa besed v fonetični prepis	27
3.3.1 Napovedovanje naglasnega mesta	28
3.3.2 Določanje fonetičnega zapisa besed	31
3.4 Preverjanje pravilnosti samodejnega določanja izgovarjave besed	38
4. Prozodične lastnosti govora	40
4.1 Naglaševanje	41
4.2 Intonacija	42
4.3 Ritem in hitrost govora	43
4.4 Glasnost	43
4.5 Splošni in jezikovno odvisni prozodični pojavi	44
5. Prozodične lastnosti slovenskega knjižnega jezika	45
5.1 Trajanje glasov v slovenskem knjižnem jeziku	45
5.2 Trajanje samoglasnikov	46
5.3 Trajanje soglasnikov	48
5.4 Osnovna frekvenca samoglasnikov	49
5.5 Besedni naglas v slovenskem knjižnem jeziku	50
5.6 Tonemski besedni naglas	51
5.6.1 Besedni naglas v nevtralni intonacijski legi	53
5.6.2 Besedni naglas pod vplivom stavčne intonacije	54
5.7 Stavčna fonetika	55
5.7.1 Členitev besedila s premori	55
5.7.2 Stavčna intonacija v slovenskem knjižnem jeziku	57
5.7.3 Register	60
5.7.4 Ritem govora	60
5.7.5 Hitrost in tempo govora	61

6. Meritve prozodičnih parametrov za slovenskega govorca	62
6.1 Meritve inherentnega trajanja in inherentne osnovne funkcije za slovenske glasove	62
6.1.1 Trajanje samoglasnikov	64
6.1.2 Osnovna frekvenca samoglasnikov	66
6.1.3 Trajanje soglasnikov	68
6.2 Meritve trajanja slovenskih glasov pri različnih hitrostih govora	70
6.2.1 Analiza parov	72
6.2.2 Normirano daljšanje in krajšanje glasov	74
6.3 Meritev tempa artikulacije pri različnih hitrostih govora	76
6.4 Merjenje značilnih intonacijskih krivulj za potek stavčne intonacije	79
7. Nastavljanje prozodičnih parametrov	81
7.1 Predpriprava fonetičnega zapisa besedila	81
7.2 Modeliranje trajanja glasov in zlogov	84
7.2.1 Trajanje samoglasnikov	87
7.2.2 Trajanje soglasnikov	87
7.2.3 Ekstrinzično trajanje taktov	88
7.2.4 Preverjanje uspešnosti napovedanega trajanja	89
7.3 Nastavljanje trajanja premorov	91
7.4 Modeliranje osnovne frekvence	91
7.4.1 Oblikovanje besednega naglasa	94
7.4.2 Oblikovanje stavčne intonacije	96
8. Sinteza govornega signala	98
8.1 Zbirka osnovnih enot	100
8.2 Združevanje osnovnih enot	102
9. Preizkus sintetizatorja govora	104
10. Zaključek	107
Uporabljeni viri in literatura	110
Terminološki slovarček	128
Povzetek	140
Automatic text-to-speech synthesis (Summary)	144

Linguistica et philologica

1999

Alenka Šivic-Dular

*Besedna družina iz korena *god- v
slovanskih jezikih. Pomenoslovna
razčlemba v kulturološkem kontekstu*

1998

Varja Cvetko Orešnik

*K metodologiji preučevanja
baltoslovansko-indoiranskih jezikovnih
odnosov. Prvi del*

ISBN 961-6358-21-9



9 789616 358217