

# GOVOREČI RAČUNALNIK

*Jerneja Gros, France Mihelič, Nikola Pavešić*

Fakulteta za elektrotehniko

Tržaška cesta 25

tel. (061) 1768 316, fax (061) 1264 900

nejka@fe.uni-lj.si

## POVZETEK

V članku opisujemo različne postopke in meritve, ki smo jih udejanili oz. izvedli pri razvoju sintetizatorja govora za slovenski jezik. Sintetizator govora smo uporabili za podajanje odgovorov v samodejnem odzivniku za poizvedovanje o letalskih informacijah.

Sintetizator govora lahko pretvori poljubno slovensko besedilo v razumljiv računalniški govor. Sprva se vhodno besedilo pretvori v zaporedje fonetičnih simbolov. Nato za vsak glas napovemo njegovo trajanje, za zveneči glas pa še višino osnovnega tona, na kateri naj bo glas izgovorjen. Sledi preoblikovanje in povezovanje kratkih, vnaprej posnetih delov govornega signala, pri čemer se upoštevata želeno trajanje in višina (zvenečih) glasov.

## ABSTRACT

A text-to-speech system, capable of synthesising continuous Slovenian speech from an arbitrary input text is presented.

The input text is transformed into its spoken equivalent by a series of modules, which we describe in detail. Further, a special approach to prosody modelling is presented.  $F_0$  modelling is based primarily on predicting the appropriate tonemic accent. Phone durations are predicted by a two-level approach, taking into account how acceleration or slowing down apply to the durations of individual phones. The TTS system is based on the concatenation of basic speech units, diphones, using the TD-PSOLA technique.

## 1 UVOD

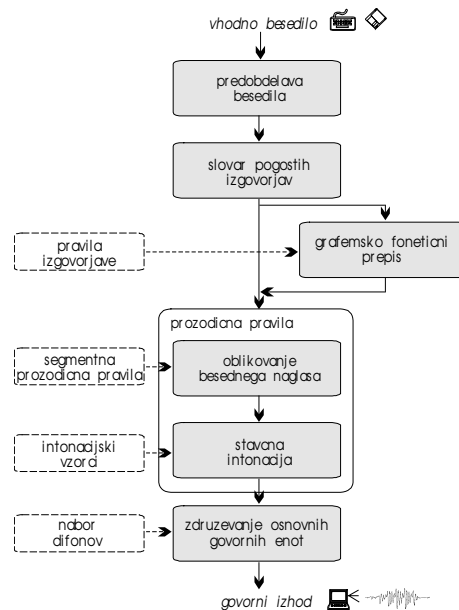
Sinteza govora je proces umetnega ustvarjanja govora, ki ga lahko uporabljamo za najrazličnejše namene. Sistemi za pretvorbo besedila v govor omogočajo posredovanje informacij, kot so teletext, elektronska pošta, faksi in druga besedila, brez potrebe po njihovi vidni predstavitvi, kar je še zlasti uporabno za slepe in slabovidne osebe. Nadalje se uporabljajo v sistemih za avtomatsko posredovanje informacij, tako imenovanih interaktivnih govornih odzivnikih, kjer bi bilo tovrstno delo za človeka enolično in utrujajoče [Sorin94].

Prvi poskus v smeri sinteze slovenskega jezika je izvedel J. Hribar [Hribar84]. 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 preprost način priredil prozodične parametre in ga posredoval dvema formantnima sintetizatorjema, razvitima za angleški jezik [Weilguny93].

Raziskave na omenjenem inštitutu nadaljujeta A. Dobnikar in T. Šef z razvojem difonskega sintetizatorja govora [Šef97]. Razvili so tudi model za napovedovanje slovenske stavčne intonacije in model za napovedovanje trajanja premorov [Dobnikar97].

Tu velja omeniti tudi govoreči računalnik, ki ga je razvil A. Šurlan, in ga je posebej prilagodil za potrebe slepih in slabovidnih.



Slika 1. Zgradba sistema za sintezo slovenskega govora.

V Laboratoriju za umetno zaznavanje na Fakulteti za elektrotehniko v Ljubljani smo pričeli z delom na razvoju sintetizatorja govora v letu 1994. Kot cilj smo opredelili razvoj sintetizatorja govora, ki naj bi bil sposoben kakovostno samodejno pretvarjati slovenska besedila v govor. Zato je bilo potrebno poskrbeti za prijetno melodijo izgovora in se čim bolj približati pravilnemu naglaševanju besed. Sintetizator govora smo uporabili za podajanje odgovorov v samodejnem odzivniku za poizvedovanje o letalskih informacijah [Ipšič97].

Sistem za sintezo slovenskega govora smo zasnovali modularno [Gros97a], glej sliko 1. Vhod v sintetizator govora predstavlja poljubno besedilo, ki je shranjeno v računalniški obliki. Besedilo se sprva pretvori v ustrezen 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 [Moulines90]. Za osnovne govorne enote smo izbrali difone. Predvajanje digitalnega zapisa govornega signala poteka preko zvočne kartice računalnika.

V nadaljevanju opisujemo delovanje posameznih modulov sistema in način pridobivanja baze difonov. Članek sklenemo z rezultati preskusa razumljivosti in naravnosti sintetiziranega govora.

## 2 DOLOČANJE IZGOVORJAVE BESEDILA

Vhod v sintetizator govora predstavlja poljubno besedilo v slovenskem jeziku, zapisano v ASCII formatu. Pretvorba besedila v fonetični prepis poteka v dveh korakih – sprva se izvede predobdelava besedila, nato sledi pretvorba grafemskega zapisa besedila v fonetični prepis.

### 2.1 Predobdelava besedila

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

Rezultat predobdelave je besedilo, zapisano kot zaporedje ločil in besed s črkovnimi grafemi.

### 2.2 Pretvorba vhodnega besedila v fonetični prepis besedila

Izgovorjavo besede najprej iščemo v slovarju pogostih izgovorjav. Do slovarja pogostih izgovorjav smo prišli s statistično analizo obsežne zbirke slovenskih besedil. Tem besedam smo ročno določili fonetični prepis izgovorjave. Slovar izgovorjav obsega več kot 16.000 najbolj pogostih besed, ki so tematsko razdeljene v šest podslovarjev.

Če besede ne najdemo v slovarju izgovorjav, ji moramo izgovorjavo določiti samodejno. Sprva napovemo naglasno mesto, nato besedo preslikamo v njen fonetični prepis.

### 2.3 Napovedovanje mesta naglasa

Mesto naglasa predstavlja zlog, na katerem ima beseda jakostno ali tonsko izrazitost [Toporišič92, str. 105]. Glede na besedno obliko poznamo stalno mesto naglasa, kot npr. v francoščini na zadnjem zlogu, delno omejeno mesto naglasa, kot npr. v hrvaščini zadnji zlog ni nikoli naglašen, ter prosto mesto naglasa.

Za slovenski jezik je značilno prosto mesto naglasa, saj se ta lahko pojavi na prvem, zadnjem, predzadnjem ali predpredzadnjem zlogu. Prav tako ima lahko posamezna beseda več mest naglasa. Mesto naglasa je določeno za vsako besedo posebej in se ga naučimo hkrati z učenjem jezika in besed.

Vendarle obstajajo pravila, ki vsaj delno določajo mesto naglasa [Toporišič91, str. 53-56]. Ta pravila smo uporabili za samodejno določanje mesta naglasa v besedi. Omenjena pravila ne zaobjamejo vseh besed. Zato smo slovar pogostih izgovorjav statistično analizirali in poiskali pravila, ki napovedujejo naglasno mesto v besedi glede na število zlogov v besedi. S temi pravili smo določili najbolj verjetno naglasno mesto v besedi.

## 2.4 Samodejna pretvorba besed v fonetični prepis

Za pretvorbo besed v njihovi izgovorjavi pripadajoča zaporedja fonetičnih simbolov smo uporabili množico preko 150 produkcijskih pravil za posamezne glasove oziroma skupine glasov. Pravila smo ponovno optimizirali na slovarju pogostih izgovorjav. Pri določanju izgovorjave besed smo upoštevali tudi koartikulacijske vplive na stikih med besedami.

## 3 NASTAVLJANJE PROZODIČNIH PARAMETROV

Pravilnost izbire prozodičnih parametrov močno vpliva na kakovost sintetiziranega govora. Postopek nastavljanja prozodičnih parametrov se deli na dva poglavitna koraka: napovedovanje trajanja za posamezni glas in napovedovanje osnovne frekvence za zvoneče glasove.

### 3.1 Modeliranje trajanja

Dejavniki in parametri, ki jih upoštevamo pri nastavljanju trajanja glasov, so skladenjske, fonetične in fiziološke narave [Rejec88, str. 23-32], [Epitropakis93a]:

- segmentna identiteta: pove ime glasu, ki mu določamo trajanje;
- fonološki: inherentno trajanje glasu
- 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 nenaglašenem zlogu
- položaj zloga v taktu: naglašeni, prednaglasni, ponaglasni, vmesni, začetni, končni
- položaj glasu znotraj besede: npr. če se beseda prične s samoglasnikom, se le-ta podaljša
- vrsta zloga: npr. ali se glas nahaja v odprtem ali v zaprtem zlogu;
- fiziološki: minimalno trajanje posameznega glasu;
- skladenjski: položaj takta znotraj fraze, tako ločimo izolirani takt, takt pred premorom, takt za premorom, zadnji takt v frazi ipd.;

- dolžina takta: podaja jo število zlogov v taktu, tako poznamo enozložne, dvožložne takte ipd.;
- hitrost govora: vpliva na trajanje takta, v katerem se glas nahaja.

Za nastavljanje trajanja govornih enot za slovenski jezik smo uporabili *dvostopenjski pristop* kontrole trajanja [Epitropakis93c] na naslednji način:

- 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 (frazе, stavke), se skrajšujejo ali podaljšujejo, v skladu z zahtevami višjenivojskih prozodičnih pojavov, in privzamejo *ekstrinzično dolžino*. Pri določanju ekstrinzične dolžine takta upoštevamo različne skladienske parametre, kot so položaj takta v frazi, izbrano hitrost govora in dolžino takta, izraženo 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 potrebno 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 preklon med obema nivojema za razliko od [Ferreira93], ki za preklonno govorno enoto jemlje besedo, izberemo *takt*. S tem se izognemo nevarnosti, da bi brezglasnice, 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 meritvami trajanj glasov v izbranih ločeno izgovorjenih 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 razvili 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 smo pravilno upoštevali različno relativno spreminjanje trajanja posameznih glasov ob spremembi trajanja celotnega takta [Gros97b].

### 3.1.1 Intrinzično trajanje glasov

Nastavljanje intrinzičnega trajanja glasov poteka na naslednji način. Sprva se nastavijo začetne vrednosti, ki so enake vrednostim, ki smo jih dobili pri merjenju trajanja in osnovne frekvence naglašanih in nenaglašanih samoglasnikov ter zvenečih soglasnikov izbranega govorca, podobno kot v [Srebot88].

Pri tem upoštevamo strukturo besede glede položaja naglašene zloga (bariton ali oksiton), pri samoglasnikih pa še dolgost ali kratkost samoglasnika, naglašenost (naglašen, nenaglašen), lego zloga (naglasni, prednaglasni, ponaglasni) ter odprtost zloga (odprt ali zaprt). Sledi popravljanje začetnih vrednosti glede na vrsto naglasa. Tako se na primer podaljšajo naglašeni zlogi z akutnim naglasom ter skrajšajo naglašeni zlogi s cirkumfleksnim naglasom.

### 3.1.2 Ekstrinzično trajanje glasov

Da bi v sintetičnem govoru vsaj delno ustvarili vtis kvziritmičnega menjavanja naglašanih in nenaglašanih 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 [Shaughnessy95], kjer so izključeni vsi premori, ne glede na njihov njihov značaj ali trajanje. Ker je vpliv naglašenosti zloga upoštevan že pri nastavljanju intrinzičnega trajanja glasov, ga pri nastavljanju ekstrinzičnega trajanja ne upoštevamo.

### 3.1.3 Usklajevanje intrinzičnega trajanja z ekstrinzičnim

Za pravilno upoštevanje rezultatov modeliranja trajanja različnih govornih enot iz obeh modelov smo razvili tri postopke, ki intrinzično dolžino takta, določeno z vsoto intrinzičnih dolžin posameznih glasov v taktu, prilagodijo zahtevani ekstrinzični dolžini, in pri tem vzpostavijo pravilna razmerja med trajanji posameznih glasov znotraj takta. Poimenovali smo jih postopek skaliranja, postopek translacije in proporcionalni postopek.

Pri dveh predlaganih postopkih upoštevamo različno relativno spreminjanje trajanja posameznih glasov ob spremembi trajanja celotnega takta [Gros97b].

Dobljene vrednosti ekstrinzičnega trajanja glasov pri normalni hitrosti govora smo primerjali s trajanji istoležnih glasov v naravnem govoru na testni bazi, ki je obsegala 172 stavkov. Povprečna vrednost absolutne razlike med trajanjem glasov pri naravnem govoru in med trajanjem glasov, napovedanim z metodo translacije, je bila le 11 ms [Gros97b].

Še boljše ujemanje naravnega in napovedanega trajanja glasov se je pokazalo pri naglašanih samoglasnikih, kjer je povprečna vrednost absolutne razlike med trajanjem

glasov pri naravnem govoru in med trajanjem glasov, napovedanim z metodo translacije, znašala le 8 ms. Standardna deviacija napovedenega trajanja od naravnega trajanja glasov je bila 15.4 ms, za naglašene samoglasnike le 13.2 ms. Po podatkih iz [Huggins71] leži tako odstopanje še pod mejo najmanjše zaznavne spremembe trajanja (angl. JND ali just noticeable difference) v tekočem govoru.

### 3.2 Modeliranje osnovne frekvence

V slovenskem knjižnem jeziku naglašujemo na dva načina, jakostno in tonemsko [Toporišič, 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 – na oblikovanje besednega naglasa ter na oblikovanje stavčne intonacije.

#### 3.2.1 Oblikovanje besednega naglasa

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
- razgibavanje osnovnih frekvenc.

Pri nastavljanju začetnih vrednosti samoglasnikov smo, v skladu s [Srebot88, 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.

#### 3.2.2 Oblikovanje stavčne intonacije

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.

Težišče intonacije določimo s pomočjo zbirke vprašalnic z označenimi mesti težišča. Če se v stavku nahaja vprašalnica iz zbirke, je težišče na tej vprašalnici, sicer težišče postavimo na naglašeni zlog zadnje besede oz. takta v frazi.

Za nastavitev stavčne intonacije smo uporabili preprost postopek, ki ne zahteva predhodne skladišne oziroma pomenske razgradnje stavka [Sorin87]. Pri določanju nastavitvenih pravil smo se naslonili na predhodne ugotovitve jezikoslovcev o stavčni intonaciji za slovenski govor [Toporišič69].

Njihove ugotovitve smo potrdili z opazovanjem poteka osnovne frekvence na množici testnih stavkov, vzeti iz dnevnega časopisja, ki jih je prebral govorec. Za določanje osnovne frekvence smo uporabili algoritem s funkcijo povprečne razlike amplitud. Določili smo značilne poteke ovojnice osnovne frekvence za štiri osnovne oblike stavčne intonacije: pripovedno, vprašalno, nekončno in vzklično.

## 4 SINTEZA GOVORA Z ZDRUŽEVANJEM OSNOVNIH GOVORNIH ENOT

Za samo sintezo govornega signala smo uporabili postopek TD-PSOLA (Time Domain Pitch-Synchronous Overlap-Add synthesis) [Moulines90], 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 zelenimi vrednostmi osnovne frekvence in časa trajanja za posamezen glas. Na izhodu modula dobimo sintetiziran govor.

### 4.1 Baza difonov

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, izgovorjenih z razmeroma enakomerno intonacijo. Besede smo izbrali tako, da se zeleni difon nahaja v fonetično čim bolj nevtralnem okolju z ostalimi glasovi.

Bazo difonov smo ročno označili: določili smo sredino prehoda med glasovoma znotraj difona, na zvonečih delih signala pa smo postavili značke osnovne frekvence. Celotno bazo smo normirali na glasnost, tako da se na mestu zleпка srečata dela signala z enako glasnostjo.

## 5 PRESKUS SINTETIZATORJA GOVORA

Da bi ugotovili kakovost, naravnost in možnost uporabe umetno sintetiziranega govora, smo v Laboratoriju za umetno zaznavanje pripravili preskus sintetizatorja govora, na 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. 18 poslušalcev pred preskusom še ni slišali sintetizatorja govora.

Celoten preskus je trajal manj kot eno uro. Preskus 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 vsaj delno navadili na sintetizirani govor.

Del preskusa smo zasnovali po priporočilih mednarodne zveze za telekomunikacije ITU [ITU93,ITU94], v preostalem delu preskusa smo želeli dobiti oceno posameznih nastavitvev sintetizatorja govora.

Preskus kakovosti sintetizatorja govora za slovenski jezik smo razdelili na tri dele:

- ocenjevanje primernosti sintetizatorja govora za uporabo v samodejnem odzivniku za posredovanje letalskih informacij,
- primerjava kakovosti sintetičnega govora s pošumljenim naravnim govorom,
- opisne ocene kakovosti sintetiziranega govora.

Prvi del preskusa 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 preskusu 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 5dB in 10dB. V tretjem delu testa so poslušalci ocenjevali različne nastavitve parametrov sintetizatorja govora.

## 6 ZAKLJUČEK

Kakovost končnega sintetiziranega govora je odvisna od vseh modulov v sistemu. Preskus 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.

Nadalje načrtujemo izvesti skladenjsko in pomensko razgradnjo povedi, ki bo pomagala pri odpravljanju dvoumnosti v izgovorjavi besed in bo omogočala bolj pravilno nastavitvev stavčne intonacije. Pričakujemo, da bo to dodatno izboljšalo kakovost sintetiziranega govora. V kratkem nameravamo tudi postaviti interaktivni strežnik, ki bo omogočal tvorjenje sintetičnega govora preko Interneta. Strežnik bo od jeseni 1998 naprej dostopen na naslovu <http://sinteza.fe.uni-lj.si>.

## 7 VIRI

[Dobnikar97] A. Dobnikar. *Določevanje stavčne intonacije pri sintezi slovenskega govora*. Doktorska disertacija. Fakulteta za elektrotehniko. 1997.

[Epitropakis93a] G. Epitropakis, D. Tambakas, N. Fakotakis, G. Kokkinakis. *Duration modelling for the Greek language*. Zbornik EUROSPEECH'93. Berlin. str. 1995-1998. 1993.

[Epitropakis93b] G. Epitropakis, N. Yiourgalis, G. Kokkinakis. *Prosody control of TTS-systems based on linguistic analysis*. Zbornik EUROSPEECH'93. Berlin. str. 1999-2002. 1993.

[Ferreira93] F. Ferreira. *Creation of prosody during sentence production*. Psychological review. (2). str. 233-253. 1993.

[Gros97a] J. Gros. *Samodejno tvorjenje govora iz besedil*. Doktorska disertacija. Fakulteta za elektrotehniko. 1997.

[Gros97b] J. Gros, N. Pavešić, F. Mihelič. *Syllable and segment duration at different speaking rates in the Slovenian language*. Zbornik EUROSPEECH'97. 1997.

[Hribar84] J. Hribar. *Sinteza umetnega govora iz teksta*. Magistrsko delo. Univerza v Ljubljani. Fakulteta za elektrotehniko. 1984.

[Huggins71] 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.

[Ipšič97] I. Ipšič, F. Mihelič, K. Pepelnjak, J. Gros, S. Dobrišek, N. Pavešić, E. Nöth, *The Slovenian dialog system for airflight inquiries*. Zbornik 2. SQEL Workshop on Multi-Lingual Information Retrieval Dialogs. Pilsen. str. 133-136. 1997.

[ITU93] ITU-T. *Recommendation P.81. Telephone transmission quality subjective opinion tests - Modulated noise reference unit*. Blue Book. (5). str. 1-5. 1993.

[ITU94] 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.

[Moulines90] E. Moulines, F. Charpentier. *Pitch-Synchronous waveform processing techniques for text-to-*

*speech synthesis using diphones*. Speech Communication. (9). str. 453-467. 1990.

[Sorin87] C. Sorin, D. Larreur, R. Llorca. *A rhythm based prosodic parser for text-to-speech systems in French*. Zbornik ICPHSc'87. Tallin. Estonija. str. 125-128. 1987.

[Sorin94] C. Sorin. *Towards High-Quality Multilingual Text-to-Speech*. Zbornik CRIM/FORWISS Workshop on Progress and Prospects of Speech Research and Technology. Uredili H. Niemann in drugi. München. str. 53-62. 1994.

[Srebot88] T. Srebot Rejec. *Word Accent and Vowel Duration in Standard Slovene: An Acoustic and Linguistic Investigation*. Slavistische Beiträge. Zvezek 226. Verlag Otto Sagner. München. 1988.

[Šef97] T. Šef, A. Dobnikar. *Izpopolnjeni alogoritem za spreminjanje osnovnega tona in časa trajanja difonov pri sintezi govora*. Zbornik ERK'97. Portorož. Zvezek B. str. 221-224. 1997.

[Toporišič69] J. Toporišič. *Slovenki jezik, izgovor i intonacija s recitacijama na pločama*. 1969.

[Toporišič91] J. Toporišič. *Slovenska slovnica*. Založba Obzorja. Maribor. 1991.

[Toporišič92] J. Toporišič. *Enciklopedija slovenskega jezika*. Cankarjeva založba. Ljubljana. 1992.

[Weilguny93] S. Weilguny. *Grafemsko-fonemski modul za sintezo izoliranih besed slovenskega jezika*. Magistrsko delo. Fakulteta za elektrotehniko in računalništvo. Univerza v Ljubljani. 1993.