

## Sistem GOVOREC za sintezo slovenskega govora

Tomaz Šef

Institut "Jožef Stefan", Jamova 39, 1000 Ljubljana, Slovenija  
E-pošta: tomaz.sef@ijs.si

**Povzetek.** Sistem Govorec za sintezo neomejenega slovenskega govora sestavlja več med seboj povezanih in hierarhično urejenih modulov: to so analiza besedila, nastavljanje prozodičnih parametrov in generiranje govornega signala. V članku so podrobneje opisani zgradba posameznih modulov ter uporabljene tehnike in metode dela. Sistem je zasnovan na obsežnih označenih tekstovnih in govornih korpusih ter je rezultat dolgoletnega raziskovalnega in aplikativnega dela Skupine za jezik in govor v okviru Odseka za inteligentne sisteme na Institutu "Jožef Stefan". Govorec je uporabljen v več aplikacijah; med drugimi ga uporabljajo tudi člani Zveze društev slepih in slabovidnih Slovenije. Trenutno sistem testira več podjetij s področja računalništva in telekomunikacij za gradnjo dinamičnih govornih portalov oz. za posredovanje raznih informacij po telefonu (e-pošta, SMS, vremenske napovedi, prometne informacije, branje poljubnih internetnih strani itd.).

**Ključne besede:** sinteza govora, naravni jezik, inteligentni sistemi, dinamični govorni portali

## Slovenian text-to-speech system GOVOREC

**Extended abstract.** While for the major world languages a rather high level of quality of artificial speech has been achieved, the Slovenian language lacks high-quality suitably-marked text and speech corpuses, dictionaries and other linguistic sources in a computer readable form. To fill-up the gap, an extensive research in the field of linguistic technologies has been recently carried out in Slovenia. The result are several new systems, some of them already used in practice. At the "Jožef Stefan" Institute we developed a system called Govorec (Speaker). It allows for an automatic conversion of any Slovenian text into speech. The different phases of the synthesis task are performed by several sequentially operating independent modules: text analysis, prosody generation and speech generation.

The first module comprises text normalization and grapheme-to-phoneme conversion tasks, such as end-of-sentence detection, abbreviation and number expansion, special format conversion, morphological analysis and phonological modelling, accentuation and phrasing. With an extra effort, the weakest point of our synthesizer, the correct lexical allotment of the stress to words, has been eliminated, too. The following fields were the ones we investigated in more detail: development of a comprehensive morphological and phonetic dictionary, optimization of algorithms for automatic accentuation of unknown Slovenian words, homograph dissambiguation.

The prosody parameters assignment involves duration, pitch and intensity modelling, and insertion of pauses. In order to generate rules for our synthesis scheme, data were collected by analysing readings of ten speakers, five males and five females. Eight of them are professional speakers of the Slovenian national radio company. Our work was based on a two-level approach for duration modelling and on the so-called superpositional approach for pitch modelling. Higher naturalness and agitation of the synthetic speech were achieved mainly through different transformations between labelled speech corpus and concrete text that we synthesised.

The whole procedure ends with speech signal generation based on the concatenation of prerecorded speech units, unit se-

lection methods and TD-PSOLA or HNM+ technique. The units inventory for the elaborated male voice (a professional speaker of the Slovenian national radio company) consists of 1509 diphones, 1568 words and 139 sentences. It was hand-segmented and hand-labelled in order to enable optimal coupling at concatenation points.

The system was first implemented in the EMA employment agent and is now used by members of the Slovenian Association of the Blind and Weak-Sighted. It was first prize awarded for valuable innovation in the field of life improvements for handicapped people and is freely accessible for non-commercial purposes through the Internet. Currently, the system is tested by several leading Slovenian telecommunication companies to determine its capacity for providing information (e-mail, SMS, weather reports, traffic information) through mobile phones.

Our experiments show that our text-to-speech (TTS) system is an appropriate tool for generating audible speech from text, written in the Slovenian language. Though the quality of the artificial speech is in general still inferior to the human speech, its improvements are already noticeable and valuable.

**Key words:** text-to-speech synthesis, natural language, intelligent systems, telecommunication applications

---

### 1 Uvod

Glede na to, da je govor eden najstarejših, najnaravnejših in v številnih situacijah najpomembnejših načinov sporazumevanja med ljudmi, ni presenetljivo, da z razvojem tehnologije postaja čedalje bolj nujno potreben tudi na različnih področjih multimedije, v telekomunikacijah, informacijskih sistemih, sistemih za opozarjanje, elektronski pošti, prevajalnih sistemih in ne nazadnje pri pomoči prizadetim osebam.

Projekcije predvidevajo občutno povečanje deleža tovrstne komunikacije v svetu. Razvoj na tem področju je za nekatere jezike dosegel že precej visoko raven [1, 2] in je rezultat načrtnih dolgoletnih raziskav in interaktivnega sodelovanja strokovnjakov različnih tehničnih in humanističnih ved. Frederik Hayes-Roth v članku "AI - What Works and What Doesn't?", objavljenem v reviji *AI Magazine* [3], uvršča današnje sisteme za generiranje umetnega govora med največje dosežke umetne inteligence. Po Patti Maes prav govorna komunikacija v inteligentnih sistemih (agenti) omogoča "humanizacijo" računalnikov [4].

Razvoj sinteze slovenskega govora je usmerjen predvsem v raziskave jezikovno odvisnih delov sistema. Pri tem delu gre za izvirne prispevke, saj je vsak govor popolnoma specifičen in s parametri težko primerljiv, vsekakor pa nenadomestljiv z drugimi jeziki oz. parametri govora. Sledenje svetovnim trendom na tem področju je pomembno tako za obstoj slovenske nacionalne identitete kakor tudi za enakopravno vključevanje v globalni elektronski prostor.

Prvi praktični rezultat dela je bil Sistem za izgovorjavo izoliranih besed slovenskega jezika, razvit na Odseku za inteligentne sisteme na *Institutu "Jožef Stefan"* v Ljubljani [5]. Delo na tem področju se je nadaljevalo z razvojem in dopolnjevanjem sintetizatorja slovenskega govora, sposobnega samodejnega pretvarjanja poljubnih slovenskih besedil v govor - sistem STTS oz. sistem GOVOREC [6, 7, 8]. Podoben sistem z oznako S5 so vzporedno začeli razvijati tudi na *Fakulteti za elektrotehniko v Ljubljani* [9,10]. Razvoj sistemov za sintezo govora v Sloveniji ponazarja slika 1. Polne puščice označujejo obdobje razvoja oz. nadgradnje posameznega sistema, črtkane puščice pa ponazarjajo prenos rezultatov nekaterih jezikoslovnih raziskav v posamezno aplikacijo oz. sočasno sodelovanje istih avtorjev pri hkratnem razvoju različnih sistemov.

V zadnjih treh letih so v sodelovanju s Centralnimi raziskovalnimi laboratoriji podjetja Siemens tudi na *Fakulteti za elektrotehniko, računalništvo in informatiko v Mariboru* začeli izdelavati večjezični sistem sinteze govora [11]. Njihovo delo je usmerjeno predvsem na razvoj postopkov in orodij za analizo besedila v procesu pretvorbe besedila v govor. Prav tako so se začeli ukvarjati z analizo čustev v govoru [12].

## 2 Zgradba sistema GOVOREC

Tako kot skoraj vsi podobni sistemi v svetu je tudi sistem GOVOREC za sintezo slovenskega govora zasnovan modularno (slika 2). V hierarhični zgradbi modulu za analizo besedila (predobdelava besedila in grafemsko fonemska pretvorba) sledi modul za nastavljanje proizvodnih parametrov (trajanje, osnovna frekvenca, amplituda, premori), temu sledi še modul za končno generiranje

govornega signala (izbira osnovne enote, lepljenje, sprememba govornih parametrov). Modularnost ni omejena le na zgornje tri kategorije, ampak se v veliki meri kaže tudi znotraj njih samih. To omogoča preprosto popravljanje in dopolnjevanje posameznih delov sistema.

Vhod v sintetizator je poljubno slovensko besedilo v digitalni obliki, shranjeno v različnih formatih. Med samim besedilom lahko nastajajo različne kontrolne oznake, ki služijo krmiljenju sintetizatorja ali pa so mu v pomoč pri doseganju večje naravnosti umetnega govora. Slednjemu služi tudi vrsta nastavljivih parametrov. Tako lahko spreminjamo način predobdelave besedila, in sicer glede na vrsto besedila, ki ga sintetiziramo (splošna besedila, kratka sporočila, matematični zapisi, časopisni članki, internet, elektronska pošta, črkovanje itd.). Pri nastavljanju proizvodnih parametrov določamo hitrost govora (normalna, hitra, počasna ali katerakoli vmesna stopnja), osnovno frekvenco, pogostnost in trajanje premorov, način naglaševanja besed, izbiramo lahko med različnimi modeli nastavljanja. V modulu za generiranje govornega signala lahko spreminjamo govorca (na voljo so trije moški glasovi), način določanja optimalnih enot lepljenja, izbiramo metode in algoritme, ki se bodo izvedli (TD-PSOLA, HNM+, linearna interpolacija, spektralno glajenje, normiranje amplitud na mestih lepljenja, kompresija, oblika izhodnega zapisa). Na izhodu dobimo sintetiziran govor, ki ga predvajamo prek zvočne kartice računalnika. Ker je sintetizator v celoti realiziran v obliki računalniškega programa, ni potrebn po dodatni strojni opreml.

Sistem GOVOREC je implementiran v skladu z industrijskim standardom Microsoft Speech API. Tako je dosegljiv najširšemu krogu uporabnikov. Zasnova sistema pa je taka, da ga je moč po potrebi prilagoditi še kateremukoli drugemu standardu.

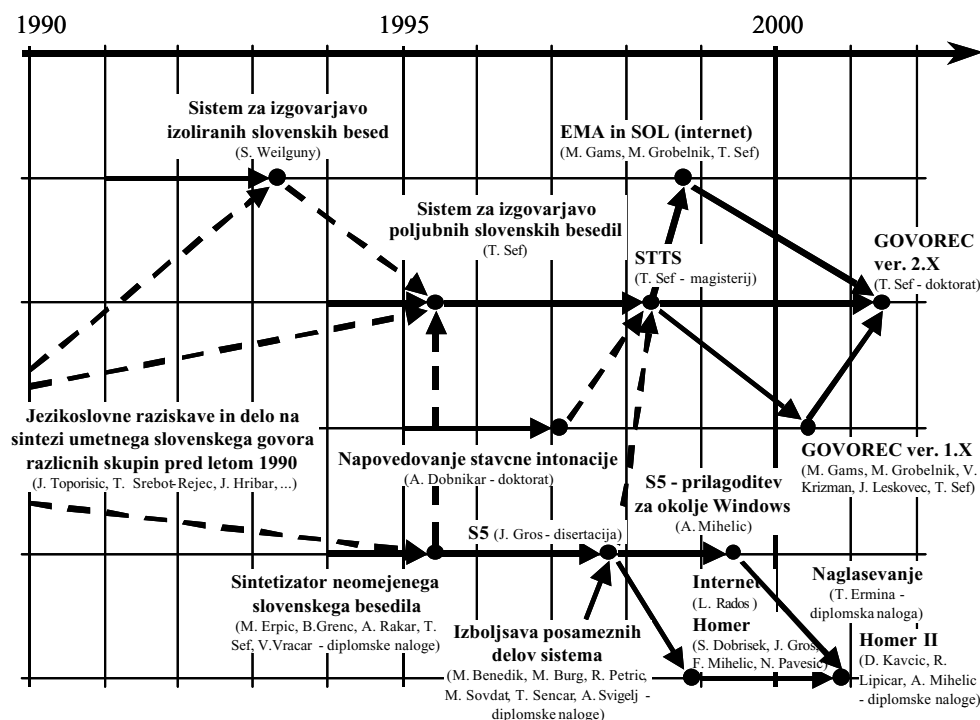
## 3 Analiza besedila

Pojem *analiza besedila* uporabljamo za vse metode in postopke, ki jih izvajamo v prvem modulu sistema GOVOREC; to je pri predobdelavi besedila in grafemsko fonemski pretvorbi. Zaradi boljšega pregleda bomo celotni opis razdelili na dva dela, in sicer glede na to, v katerem delu se posamezni postopek izvaja.

### 3.1 Predobdelava besedila

Najprej različne vhodne formate besedil pretvorimo v standardni zapis in odstranimo vse odvečne simbole, ki ne vplivajo na izgovorjavo besedila. Temu sledijo:

- določitev mej med posameznimi stavki,
- detekcija okrajšav in posebnih zapisov,
- lematizacija (gesljenje): določanje besednih enot za gesla v slovarju,



Slika 1. Ponazoritev razvoja sistemov za sintezo govora v Sloveniji  
Figure 1. History of the Slovenian TTS systems development

- določanje oblikoslovnih podatkov posameznim besedam,
- zapis v začetku detektiranih okrajšav v polni obliki,
- razvoj posebnih zapisov v ustrezen grafemski zapis.

Vhodno besedilo obdelujemo poved za povedjo. Pri tem štejemo za poved vsako zaporedje grafemov, ki se konča s končnim ločilom (pika v skladenjski rabi, vprašaj, klicaj, tri pike, EOF - poseben znak za konec datoteke). Pika nima vedno vloge končnega ločila. Ločimo naslednje neskladenjske funkcije pike: konec okrajšave (pred piko je zaporedje grafemov, ki se nahaja v seznamu okrajšav), decimalna pika, vrstilni števniki, datum.

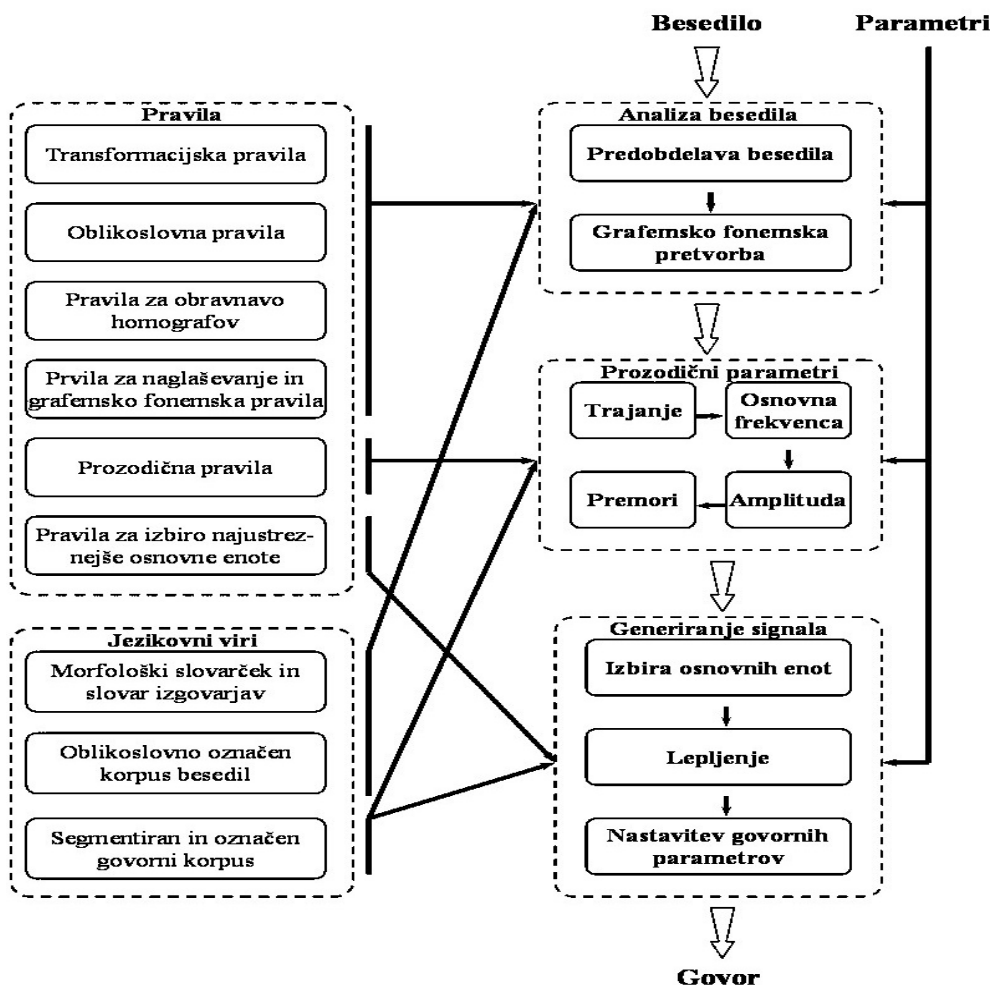
*Posebne zapise* določimo na podlagi seznama vzorcev, ki posamezni zapis opredeljujejo. Do njih smo prišli na podlagi analize obsežne zbirke primerov. Seznane nenehno dopolnjujemo in dodajamo nove. Trenutno ločimo med naslednjimi posebnimi zapisi: deljenje besed, začetki in konci odstavkov, naslovi, poudarjene besede, kratice oz. akronimi (npr. SIT za slovenski\_tolar), okrajšave besed (npr. g. za gospod), posebni simboli oz. ideogrami (npr. % za odstotek), merske in druge enote, datumi, ure, matematični zapisi, rimska števila, števniki, kolokacije. Posamezne seznane vzorcev lahko optimaliziramo tudi glede na področje uporabe sintetizatorja. Tako se npr. seznami vzorcev za sporočila, novice in podobno razlikujejo od tistih za branje matematičnih zapisov, e-pošte itd.

Rezultat normalizacije je besedilo, razdeljeno na posamezne manjše enote:

- posebni zapisi: določimo tip zapisa, pripišemo mu ustrežno oblikoslovno oznako in shranimo navodilo za njegovo nadaljnjo obravnavo;
- znane besede: iz slovarja prepisemo vse mogoče oblikoslovne oznake in ustrezne izgovarjave ter jih razvrstimo v ustrezne kategorije (npr. navadna beseda, lastno ime, del naslova, glavni števniki, morebiten pomen besede itd.);
- neznane besede (ne nahajajo se v morfološkem slovarčku in slovarju izgovarjav): glede na sosednje besede poiščemo vse potencialno mogoče oblikoslovne oznake;
- ločila v skladenjski funkciji.

V tem koraku obdelamo tudi posebna zaporedja znakov, ki služijo za krmiljenje samega sintetizatorja. Zapis najprej dekodiramo, da izluščimo njihovo informacijo (npr. hitrejša izgovarjava, višji ton, premor), ki jo nato posredujemo na ustrezna mesta znotraj posameznih enot.

Pri zapisu okrajšav v polni obliki in pri razvoju posebnih zapisov v ustrezen grafemski zapis, kakor tudi pri grafemsko fonemski pretvorbi, si v veliki meri pomagamo z oblikoslovnimi podatki besed, ki so v povedi. V pomoč so nam še razna pravila za obravnavo homografov. Pri



Slika 2. Zgradba sistema GOVOREC za sintezo slovenskega govora  
Figure 2. Architecture of the Slovenian text-to-speech system GOVOREC

tem ni pomembno, da je beseda le pravilno zapisana, ampak mora biti pravilna tudi v kontekstu.

Sestavni del našega algoritma je tako tudi *oblikoslovni označevalnik*. Za osnovo smo, podobno kot P. Jakopin [13], uporabili statistični pristop, ki se opira na slovar besednih oblik z vsemi mogočimi oznakami, slovar pogostejših besednih n-terčkov ( $n=2-5$ ; frekvenca je vsaj 2) z oznakami in slovar n-terčkov oznak ( $n=2-5$ ). N-terčki vsebujejo tudi frekvence, kar omogoča, da med postopkom spremljamo, kolikokrat se je vsaka izmed mogočih oznak pri opazovanju besedi pojavila, in seštevamo vmesne frekvence te oznake. Tako lahko v preostalih dvoumnih primerih med preostalimi oznakami izberemo najpogostejšo. Vse skupaj temelji na obsežnih oblikoslovno označenih tekstovnih korpusih besedil, ki jih prispeva podjetje AMEBIS; poleg tega prispevajo tudi modul za stavčno analizo in morfološki slovar z okoli 70.000 lemmami (700.000 različnih besednih oblik oz. 2.000.000 vnosov - ob upoštevanju vseh oblikoslovnih oznak posamezne besedne oblike).

### 3.2 Grafemsko fonemska pretvorba

Postopek poteka v več korakih in obsega:

- naglaševanje neznanih besed (tistih, ki jih ni v slovarju izgovorjav),
- detekcijo in analizo homografov (besed z enako pisno podobo in različnimi izgovorjavami),
- določitev izgovorjave besed in imen.

V sistemu GOVOREC naglaševanje besed v osnovi temelji na slovarsko podprti analizi besedila. Pri tem uporabljamo prej omenjeni morfološki slovar s 70.000 lemmami, ki smo ga s pomočjo v ta namen razvitega algoritma [8] dopolnili še s podatki o naglasu in posebnostih v izgovorjavi nekaterih besed. Algoritem (okoli 50.000 vrstic programske kode in pripadajoči slovar) omogoča pravilno izgovorjavo skoraj 300.000 lem, kar pomeni več milijonov različnih besednih oblik. To je več, kot pa jih zmora povprečen Slovenec. Tako najdemo več kot 99 odstotkov vseh iskanih besed (brez upoštevanja imen).

Za vse besede, ki so v slovarju, prepišemo mesto in vrsto tako dinamičnega kot tonemskega naglasa. Preostale besede naglasimo v dveh korakih. Najprej za vsak samoglasnik in soglasnik 'r' v besedi preverimo, če je naglašen; za naglašene glasove določimo tudi tip naglasa. Sledi popravljane tako dobljenih rezultatov glede na število naglasov v besedi in dolžino besede. Uporabimo avtomatsko generirana pravila, do katerih smo prišli z analizo fonetičnega slovarja. Pri učenju pravil smo upoštevali 67 atributov, na podlagi katerih učne vzorce razvrščamo v ustrezne razrede (vrsta naglasa, tip naglasa, prisotnost naglasa, širok ali ozek samoglasnik). Uporabljen je bil Quinlanov algoritem *See5/C5.0* [14]. Tako lahko pravilno naglasimo nepoznano slovensko besedo z več kot 83 odstotno natančnostjo. Podrobnejši opis uporabljenih postopkov in potek testiranja je podan v [15].

Odperto ostaja le še vprašanje naglaševanja raznih imen. Naš slovar do sedaj obsega nekaj tisoč takih lem (oziroma nekaj deset tisoč različnih besednih oblik), ki jih sproti dopolnjujemo, predvsem s pogostejšimi imeni iz dnevnega časopisja in preostale literature.

Več izboljšav je bilo narejenih tudi pri združevanju besed v daljše enote, kjer naletimo na problem homografije. Izkazalo se je, da lahko v povprečju vsako deseto besedo obravnavamo kot homograf. Na srečo je ponavadi ena izgovarjava veliko verjetnejša od preostalih, kar zmanjšuje verjetnost napake. S pomočjo slovarja izgovarjav smo določili nekaj tisoč homografov, ki smo jih nato razvrstili v skupine (glede na lemo in oblikoslovno oznako), iz obsežne zbirke primerov (2.000.000 besed) pa smo izluščili pravila za njihovo obravnavo. Ta upoštevajo tudi osnovni pomen nekaterih besed. Predstavljena pravila [8] dobro rešujejo predvsem najpogostejše primere in s tem precej zmanjšajo količino napak.

Pri določitvi izgovarjave besed in imen uporabljamo sezname kontekstno odvisnih grafemsko fonemskih pravil. Ta zajemajo tudi pojave, ki nastopajo pri glasovih zaradi soseščine drugih glasov. V primerih, ko poznamo oblikoslovne lastnosti besede, uporabimo specifična pravila, ki veljajo le za posamezno besedno vrsto [8]. Za druge besede pa so pravila bolj splošna.

#### 4 Nastavljanje prozodičnih parametrov

Nastavitev prozodičnih parametrov obsega določitev časa trajanja, osnovne frekvence, glasnosti in trajanja premorov. Večjo naravnost in razgibanost umetnega govora dosežemo predvsem z raznimi preslikavami (pri času trajanja) med označenim govornim korpusom in konkretnim besedilom, ki ga sintetiziramo.

*Trajanje glasov* nastavimo v dveh korakih: najprej za vsak glas določimo njegovo osnovno trajanje (na podlagi inherentnega trajanja glasu, njegovega glasovnega okolja, položaja glasu v besedi in strukture zloga, v katerem je

glas), ki ga v nadaljevanju popravimo glede na želeno hitrost govora, število zlogov v besedi in položaj besede v frazi. Vrednosti trajanja glasov avtomatično izluščimo iz označenega govornega korpusa posameznega govorca. Če je ta premalo obsežen, uporabimo predhodno izmerjene vrednosti, ki jih ustrezno prilagodimo.

*Osnovno frekvenco* glasov nastavimo s "superpozicijskim" modelom, ki potek osnovne frekvence v intonacijskem segmentu definira kot vsoto globalne komponente in lokalnih komponent [16]. Globalna komponenta definira osnovno raven poteka v celem intonacijskem segmentu, lokalne komponente pa pomenijo lokalna naraščanja in padanja pri poudarjenih delih besedila.

*Glasnost govora* lahko v sintetizatorju poljubno nastavljamo. Poleg tega se glasnost govora spreminja še glede na stavčne poudarke v besedilu.

Glede na dolžino premorov in na položaj v besedilu ločimo štiri skupine premorov [16]: premori pri naslovih, premori na koncu povedi, premori v povedi na mestih ločil in premori v povedi na mestih ritmičnih delitev (pred vezniki in, pa, ter, ali,...) v daljših stavkih, ki jih premor razdeli na krajše intonacijske segmente.

#### 5 Generiranje govornega signala

Sam postopek sinteze govornega signala temelji na korpusni sintezi govora. Govorna zbirka pri najbolj dodelanem moškem glasu (poklicni govorci s slovenskega nacionalnega radia) obsega 1509 difonov, 1568 besed in 139 skrbno izbranih in fonetično uravnoveženih stavkov. Velikost potrebnega pomnilnika je zmanjšana z različnimi postopki stiskanja govornega signala.

Predlagani algoritem za izbiro najustreznejših enot deluje na ravni povedi, pri čemer upošteva različne parametre, kot npr. spektralne karakteristike na mestih spajanja, dolžino predlagane osnovne enote, njene prozodične parametre ter vrsto in tip fonemov na mestih spajanja. Trajanje in osnovno frekvenco glasov po potrebi dodatno popravimo s pomočjo algoritma TD-PSOLA [17] ali HNM+ [18].

Trenutno lahko izbiramo med tremi moškimi glasovi, dodati pa nameravamo še ženski in otroški glas.

#### 6 Sklep

Govorec je bil najprej uporabljen v zaposlovalnem agentu EMA [19], ki posreduje informacije o prostih delovnih mestih v Sloveniji. Med drugimi ga uporabljajo tudi člani Zveze društev slepih in slabovidnih Slovenije, ki jim je bil sistem predan v brezplačno uporabo. Sistem je bil nagrajen s prvo nagrado Sklada za nagrajevanje inovacij na področju usposabljanja, življenja in dela invalidov. Stare verzije Govorca (v nekoliko okrnjeni obliki; difonska sinteza, brez slovarjev, samo en glas) so dostopne tudi prek interneta na naslovu <http://ai.ijs.si/Govorec/> in

se lahko uporabljajo v nekomercialne namene. Sintetizator trenutno testira več podjetij s področja računalništva ter telekomunikacij in sicer za gradnjo dinamičnih govornih portalov oz. za posredovanje raznih informacij po telefonu (e-pošta, SMS, vremenske napovedi, prometne informacije, novice, branje poljubnih internetnih strani itd.)

Na Institutu "Jožef Stefan" smo v sodelovanju z MID in podjetjem CDE razvili sistem GIVE, ki po telefonu (stacionarnem ali mobilnem) govori poljubno informacijo z interneta (HTML, XML, posredovanje e-pošte) oz. poljubno informacijo, ki jo uporabnik v elektronski tekstovni obliki vstavi v sistem. Nastavitev sistema je mogoča s pomočjo tonskega izbiranja, SMS ali prek računalnika; kasneje bomo vgradili še možnost prepoznavanja omejenega nabora besed. Sistem je zastavljen dolgoročno. Dodane bodo še nekatere storitve e-poslovanja, recimo nakupovanje, naročanje, pošiljanje dokumentov na faks itd. Glavna prednost je, da je sistem odprt, brezplačen in uporabnikom omogoča vnos njihovih lastnih storitev.

Podoben sistem razvijamo tudi v sodelovanju z Mestno občino Ljubljana (MOL). Njegov poglavitni namen je omogočiti bistveno boljše, bolj naravno in sprotno informiranje občanov o Ljubljani, o MOL, o vseh ljubljanskih institucijah in o dogodkih.

Direktna primerjava posameznih sistemov za sintezo slovenskega govora (po vzoru sintetizatorjev za angleški jezik), ki edina omogoča, da si potencialni uporabnik ustvari neodvisno mnenje o njihovi kakovosti, za zdaj ni mogoča (edini prosto dostopen sistem je stara verzija GOVORCA). Po subjektivnem mnenju zainteresiranih poslušalcev nova verzija sistema Govorec omogoča tvorbo bistveno bolj razumljivega in naravnega umetnega govora kakor dosednji sistemi. S tem naš sintetizator postaja primeren za čedalje širši krog uporabnikov. Končna kakovost se sicer še ne more primerjati s človeškim govorom, vendar pa je napredek več kot opazen.

## 7 Literatura

- [1] R. Sproat, *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, Kluwer Academic Publishers, 1998.
- [2] M. Beutnagel, A. Conkie, J. Schroeter, Y. Stylianou, A. Syrdal, *The AT&T Next-Gen TTS System*, 137th Acoustical Society of America meeting, 1999.
- [3] F. Hayes-Roth, *Artificial Intelligence - What Works and What Doesn't*, *AI Magazine*, str. 99-113, poletje 1997.
- [4] P. Maes, *Software Agents: Humanising the Global Computer*, *IEEE Internet Computing*, št. 1, str. 10-20, 1997.
- [5] S. Weilguny, *Grafemsko fonemski modul za sintezo slovenskega jezika, magistrsko delo*, Fakulteta za elektrotehniko in računalništvo, Univerza v Ljubljani, 1993.
- [6] A. Dobnikar, J. Bakran, *A New Approach for Slovene Text-to-Speech Synthesis*, *Proceedings of Microcomputers in Intelligent Information Systems MIS, MIPRO'95*, str. 312-318, 1995.
- [7] T. Šef, A. Dobnikar, *Recent Improvements in Slovene Text-to-Speech System*, *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, zvezek V, str. 2027-2030, 1998.
- [8] T. Šef, *Analiza besedila v postopku sinteze slovenskega govora, doktorska disertacija*, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2001.
- [9] J. Gros, N. Pavešić, S. Dobrišek, M. Erpič, B. Grenc, A. Rakar, T. Šef, V. Vračar, F. Mihelič, *Sistem za sintezo slovenskega govora, Zbornik četrte Elektrotehniške in računalniške konference ERK'95*, zvezek B, str. 265-268, 1995.
- [10] J. Gros, *Samodejno tvorjenje govora iz besedil, doktorska disertacija*, Fakulteta za elektrotehniko, Univerza v Ljubljani, 1997.
- [11] M. Rojc, J. Stergar, R. Wilhelm, H. U. Hain, M. Holzapfel, B. Horvat, *A multilingual text processing engine for the papageno text-to-speech synthesis system*, *6th European Conference on Speech Communication and Technology*, str. 2107-2110, 1999.
- [12] V. Hozjan, Z. Kačič, B. Horvat, *Analiza prozodičnih značilk govora za modeliranje emocij*, *Elektrotehniški vestnik*, letnik 68, št. 4, str. 213-218, 2001.
- [13] P. Jakopin, *Zgornja meja entropije pri leposlovnih besedilih v slovenskem jeziku, doktorska disertacija*, Fakulteta za elektrotehniko, Univerza v Ljubljani, 1999.
- [14] See5 system:  
<http://www.rulequest.com/see5-info.html>.
- [15] M. Škrjanc, T. Šef, M. Gams, *Using Decision Trees for Accentuation in the Slovenian Language*, *15th European Conference on Artificial Intelligence (ECAI'2002) - STAIRS'02 Symposium*, 2002 (sprejeto v objavo).
- [16] A. Dobnikar, *Določevanje stavčne intonacije pri sintezi slovenskega govora, doktorska disertacija*, Fakulteta za elektrotehniko, Univerza v Ljubljani, 1997.
- [17] E. Moulines, F. Charpentier, *Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones*, *Speech Communications* (9), str. 453-467, 1990.
- [18] Y. Stylianou, J. Laroche, E. Moulines, *High-Quality Speech Modification based on a Harmonic + Noise Model*, *Proceedings of EUROSPEECH'95*, str. 451-454, Madrid, 1995.
- [19] M. Gams, T. Šef, *A Speech Module in an Agent System*, *Engineering Intelligent Systems*, zvezek 8, št. 4, str. 225-232, CRL Publishing Ltd., 2000.

**Tomaž Šef** je diplomiral leta 1995 na Fakulteti za elektrotehniko in računalništvo v Ljubljani. Leta 1998 je magistriral, leta 2001 pa doktoriral na Fakulteti za računalništvo in informatiko v Ljubljani. Zaposlen je na Odseku za inteligentne sisteme na Institutu "Jožef Stefan", kjer se ukvarja s komunikacijo človek-stroj, procesiranjem naravnega jezika, sintezo govora ter zajemanjem tekstovnih in govornih baz podatkov.