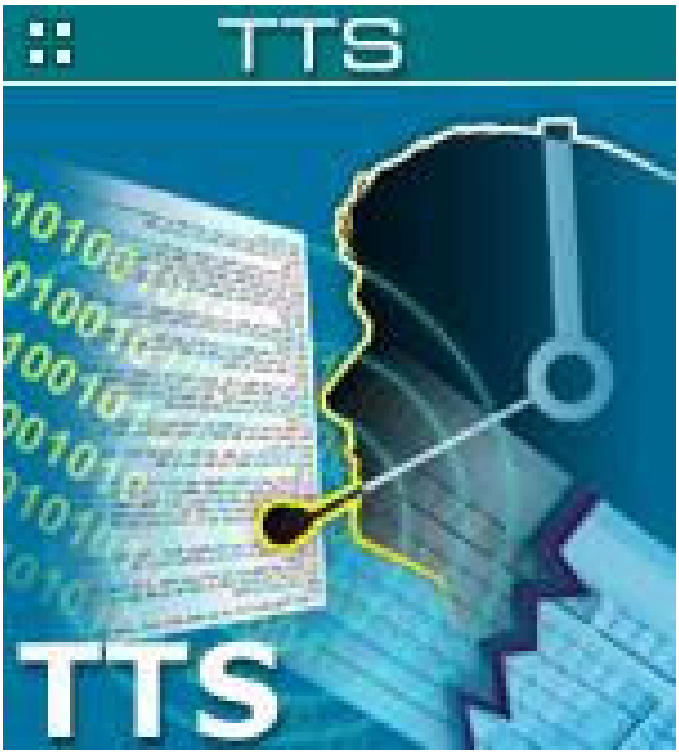
Univerza v Ljubljani
Naravoslovnotehniška fakulteta
Oddelek za tekstilstvo
Katedra za informacijsko in grafično tehnologijo

SISTEMI ZA SINTEZO GOVORA

Seminarska naloga pri predmetu JEZIKOVNE TEHNOLOGIJE



Mojca FRIŠKOVEC
Saša JANČAR

Ljubljana, maj 2006

Kazalo vsebine

Kazalo slik

Kazalo tabel

1 Uvod

Pri multimodalni komunikaciji ima govor kot komunikacijska modalnost pomembno vlogo, saj je govor človeku najbolj naravni način sporazumevanja. Razvoj sistemov, ki bodo sposobni izvajati govorno komunikacijo s človekom, zahteva razvoj na področju govornih in jezikovnih tehnologij. Govorna tehnologija vključuje sisteme avtomatskega razpoznavanja govora, sisteme avtomatske sinteze govora in sisteme govornega dialoga. Področje jezikovnih tehnologij pa vključuje črkovalnike, sisteme povzemanja besedil, sisteme tvorjenja besedil in sisteme prevajanja besedil. Slovenski jezik predstavlja zaradi svojih specifičnih lastnosti v prostoru ostalih evropskih jezikov relativno trd oreh. Največji problem povzročata pregibanje besed in nedoločen vrstni red besed v stavku. Govorna tehnologija žal še danes ni 100% zanesljiva tehnologija.

V najini seminarski nalogi bova predstavili eno od področji govorne tehnologije, to je sintezo govora.

Pri sistemih za sintezo govora lahko govorimo o dveh vrstah sistemov. Prva vrsta sistemov so sistemi s pravo sintezo govora, druga vrsta sistemov pa so sistemi z že vnaprej posnetimi glasovnimi vzorci, ki se združujejo v celoto. Slednjih sistemov ne moremo šteti med prave sisteme za sintezo govora, vendar je njihova uporaba velikokrat ekonomsko bolj upravičena kot pa izdelava pravega sistema za sintezo govora. Taki sistemi se uporabljajo tam, kjer so vsi stavki že vnaprej znani in jih je mogoče vnaprej posneti.

S stališča procesiranja signalov so ti sistemi nezanimivi. Za zanimive se izkažejo sistemi s "pravo" sintezo govora. Taki sistemi se uporabljajo kadar vsi stavki niso v naprej poznani, tako da bi jih lahko posneli v naprej.

Prvi pogoja za uspešno pretvorbo besedila v govor je izgradnja ustrezne baze izgovorjav in algoritmov. Pod sintezo govora uvrščamo predvsem tiste rešitve, kjer ni vnaprej posnetih celih besed ali celo stavkov, ampak se govor generira sproti, z lepljenjem osnovnih glasov. Nekatere rešitve, ki tačas prevladujejo, govor tvorijo z lepljenjem sestavljenih glasov, ker so rezultati glede na vložek najbolj optimalni. Obstajajo tudi druge rešitve, ki nap. simulirajo govorni trakt in dajejo zelo dobre končne rezultate. Žal je časovno krmiljenje takih sistemov izjemno zahtevno, razvoj pa izjemno dolgotrajen ter drag. Vsekakor pa se sinteza govora iz laboratorijev počasi vse bolj seli k uporabnikom, z njeno pomočjo pa številne naprave postajajo vedno bolj prijazne za uporabo, kar je glavni cilj jezikovnih tehnologij.

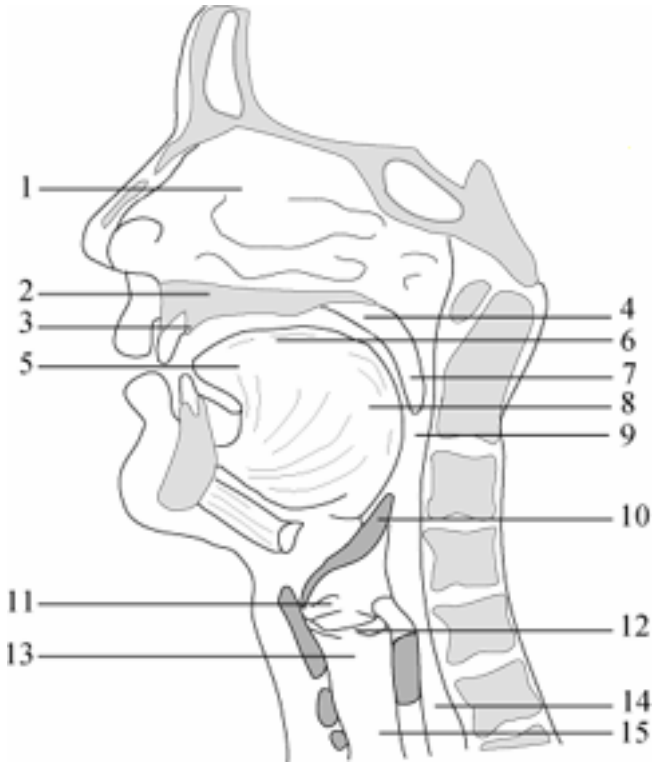
2 Govorne tehnologije

Govorne tehnologije so zelo poznano področje jezikovnih tehnologij. V grobem jih razdelimo na razpoznavo in sintezo govora. Vsaka ima svoje razvojne probleme in možnosti, obema pa je cilj vzpostaviti govorno komunikacijo med ljudmi in različnimi napravami. S pomočjo razvitih modulov že delujejo nekatere splošno znane rešitve, kot so spletni agenti in informacijski portali.

Pretvorba besedila v govor je bila na začetkih računalništva le tiha želja razvijalcev, danes pa je že povsem običajna funkcija marsikaterega sistema ali naprave. Glavni razlog za hiter razvoj tega področja v zadnjih letih je v zadostni procesorski in pomnilniški moči sodobnih naprav, ki so sposobni hraniti velikansko število informacij. Naravni jezik je namreč izjemno velika zakladnica besed, zapletenih pravil in znanja, ki ga je človeštvo ustvarjalo tisoče let. Najpomembnejši člen govorne tehnologije je prav gotovo razumevanje človekovega govornega sistema.

2.1 Človekov govorni sistem

Človek tvori govor s pomočjo govornih organov, ki so prikazani na spodnji sliki. Glavni vir energije so pljuča in trebušna prepona. Ko ljudje govorimo, zračni tok potuje skozi glasilke in grlo v tri glavne votline govornega trakta: žrelo, ustno in nosno votlino. Iz ustne in nosne votline zrak izstopi skozi nos in usta.



1 - nosna votlina
2 - trdo nebo
3 - dlesen
4 - mehko nebo
5 - konica jezika
6 - hrbtna stran jezika
7 - mehkonebni jeziček
8 - korenina jezika
9 - žrelo
10 - poklopec
11 - lažne glasilke
12 - prave glasilke
13 - grlo
14 - požiralnik
15 - sapnik

Slika 1 - Človekov govorni sistem

Prostor v obliki črke V med glasilkami je najpomembnejši vir zvoka v zvočnem sistemu. Med govorom glasilke delujejo v različnih načinih. Najpomembnejša funkcija je uravnavanje zračnega toka s hitrim zapiranjem in odpiranjem. Osnovna frekvenca vibracij je odvisna od mase in je okoli 100 Hz pri moških, 200 Hz pri ženskah in 300 Hz pri otrocih.

Žrelo povezuje grlo z ustno votlino. Ima delno spremenljive dimenzije; dolžina grla se lahko delno spreminja z zniževanjem ali zviševanjem grla na eni strani in neba na drugi strani. Mehko nebo izolira ali povezuje pot iz nosne votline do žrela. Na spodnjem delu žrela se nahaja poklopec, ki preprečuje, da bi hrana vstopala v grlo. Med požiranjem so poklopec in glasilke zapri, med normalnim dihanjem pa so odprti.

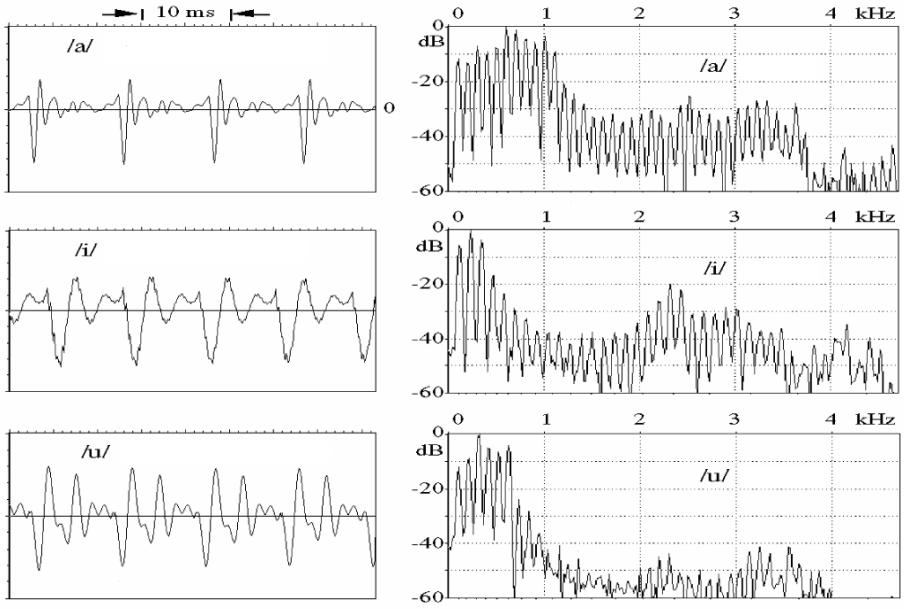
Ustna votlina je eden pomembnejših delov glasovnega trakta. Njena velikost, oblika in akustičnost se lahko spreminjajo s pomočjo premikanja neba, jezika, ustnic, ličnic in zob. Jezik je še posebno fleksibilne; konico in robove jezika lahko neodvisno premikamo, prav tako lahko celoten jezik premikamo naprej, nazaj, gor in dol. Usta uravnavajo velikost in obliko ustne odprtine skozi katero se širi zvok.

Nosna votlina ima nespremenljivo velikost in obliko. Dolga je približno 12 cm in ima prostornino 60 cm^3. Zračni tok v nosno votlino je kontroliran s pomočjo mehkega neba.

2.1.1 Govorni signal

Govor je zaporedje medsebojno odvisnih fonemov in vsebuje več informacij kot samo besedilo, saj lahko iz govorjene besede razpoznamo človekova čustva, odnos govorca in njegovo razpoloženje.

Pri predstavitvi govornega signala je najpomembnejša fonetika. To je zvočni podoba jezika od glasu, naglasa v besedi do celotnega besedila. Pomembno vlogo ima tudi fonem, ta predstavlja glas, ki nasproti drugim glasovom razločuje pomen besed. Same glasove delimo na zveneče in nezveneče. Zveneče glasove tvorijo osnovna frekvenca in višje frekvence ob pomoči glasilk. Pri nezvenečih glasovih pa so glasilke razmaknjene zato ni osnovne frekvence, niti višjih frekvenc. Glas se tvori nekje na poti do ust.



Slika 2 - Časovna in spektralna predstavitev samoglasnikov a, i, u

2.2 Področja uporabe sinteze govora

Sinteza govora, imenovan tudi tekst-v-govor (TTS) je proces kjer pretvorimo tekst v govor in je razširjen na različnih področjih. Komercialno so najpomembnejši sistemi, ki nudijo pomoč ali asistenco uporabnikom. Na primer ponoči ali ob vikendih, ko na delovnih mestih ni ljudi ali pa jih je malo, lahko uporabniki še vedno dobijo informacije preko računalnika, kot primer stanje na računu. Računalnik v takem primeru posluša govor ali pa prepozna ton pritisnjene tipke in tako odgovori s pomočjo TTS sistema. Na tak način deluje tudi polnjenje Mobi računa za mobi uporabnike. Tipično so to aplikacije, namenjene upravljanju preko telefona, čeprav jih lahko srečamo tudi v kiosih in avtomatskih govorcev.

TTS sistemi se uporabljajo za avtomobilske navigacijske sisteme, saj voznik nima časa brati zaslona, ker mora paziti na dogajanje na cestišču. Tako je za voznika enostavneje in seveda varneje, če navigacijski sistem samo posluša.

TTS se uporablja tudi na osebnih napravah, npr. na osebnem računalniku za preverjanje napisanega teksta ali za učenje novih jezikov. Sem so vključene tudi tehnologije, ki so v pomoč uporabnikom, ki so slabovidni ali slepi, in tehnologije, ki so v pomoč tistim, ki ne morejo govoriti.

2.3 Zgradba sistema za sintezo govora

Tu gre za deljenje besed v foneme ter analiza za posebno obdelavo teksta kot so: številk, valut, sklanjatev in postavljanje ločil ter generiranje digitalne avdio oblika za predvajanje.

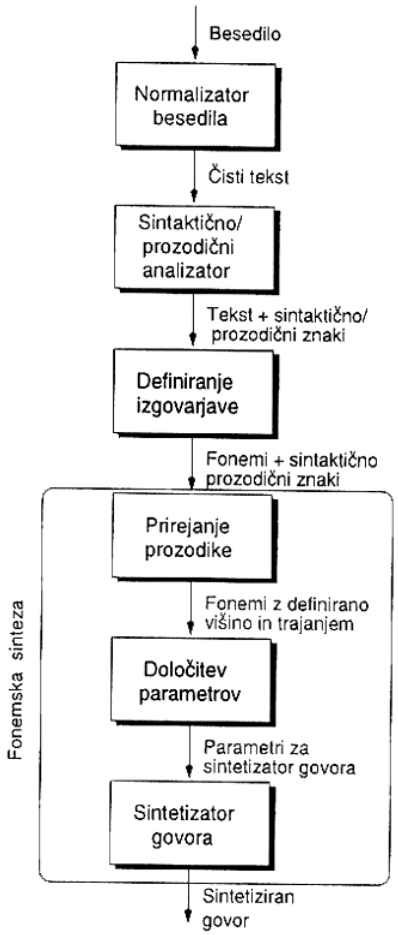
Sintetizatorji ali tekst v govor glasovi izvajajo jezikovne sinteze, rokujejo z kompleksnostjo pretvorbe teksta in generirajo govorni jezik. Ti sistemi generirajo zvoke podobne tistim, ki jih

govorijo ljudje in vključujejo različne filtre, ki simulirajo dolžino vratu, velikost ustne votline, obliko ustnic in pozicijo jezika.

Pri sami sintezi govora lahko ločimo tri glavna področja, ki so pomembna za uspešno pretvorbo besedila v govor.

Ta tri področja so:

1) Pretvorba teksta v enote sinteze
2) Pretvorba teksta v metrične parametre
3) Pretvorba sinteznih enot v govor



Slika 3 - Shema sistema pretvorbe besedila v govor

2.3.1 Pretvorba teksta v enote sinteze

Večina metod sinteze deluje na osnovi fonemov, zato so enote sinteze fonemi. Pretvorba teksta v enote sinteze je večstopenjska operacija.

Prva stopnja pri pretvorbi je normalizacija. Z normalizacijo interpretiramo odstavke, ločila. Zaporedja števk se razvijejo v ustrezne številke - glavne ali vrstilne. Okrajšave in akronimi se zapišejo v polni obliki. Zaporedje besed se nato pretvori v zaporedje enot sinteze, to so v večini

primerov fonemi. Pri pretvarjanju besed lahko najprej iščemo v slovarju pogostih izgovorjav, ki ga je predhodno potrebno zgraditi za vsak jezik posebej. Če besede ne najdemo v slovarju ali ga nimamo, ji moramo izgovorjavo določiti sami.

Drugi del pretvorbe predstavlja napovedovanje mesta naglasa. Mesto naglasa je zlog, na katerem ima beseda jakostno in tonsko izrazitost. 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.

Zadnja faza je pretvorba besed v fonetični zapis. Za to pretvorbo lahko uporabljamo produkcijska pravila za posamezne glasove, ki jih moramo predhodno sami določiti in so odvisna od jezika.

2.3.2 Pretvorba teksta v metrične parametre

Tudi določaje metričnih parametrov lahko razdelimo na tri dele.

Prvi del je določanje intonacijskih mej. Za idealno določanje mej je potrebno poznavanje celotne sintaktične analize stavka, kar pa je težko izvesti avtomatično. Pogosto pomanjkanje intonacijskih mej povzroči, da je glas bolj grob. Preveliko število intonacijskih mej pa povzročajo zmedo in motnje. Najenostavnejša metoda je postavljanje mej, kjer le te določajo ločila. Malo bolj izpopolnjena rešitev je rešitev v kateri imamo shranjeno listo funkcijskih besed pri katerih nastopajo meje. Poleg teh dveh enostavnih metod obstaja še veliko bolj izpopolnjenih metod.

Drugi del je določanje dolžin posameznih segmentov. Dolžina segmentov se ponavadi določa z množico pravil. Pravila so sestavljena iz velikega števila faktorjev, ki bi lahko vplivali na dolžino segmenta. Nekateri izmed teh faktorjev so: frekvenca besede, sintaktična kategorija, fonetičen kontekst in še mnogi drugi.

Tretji del je določanje frekvenčnih skic. Osnovna frekvenca človeškega govora je določena kot kombinacija velikega števila faktorjev. Sistemi za sintezo govora ne upoštevajo vsebine teksta, ki ga sintezirajo, ker bi morali poznati vsebino teksta. Prav tako ne upoštevajo nekaterih drugih lastnosti človeškega glasu, zato je govor bolj zglajen kot človeški.

2.3.3 Pretvorba sinteznih enot v govor

Ko imamo določene vse parametre - stavčno intonacijo, naglašene zloge besed, znakovni zapis pretvorjen v zaporedje glasov izgovorjave moramo glasove še pravilno realizirati. To lahko naredimo na dva načina:

1. Z generiranjem glasov s pomočjo simulacije človeškega govornega trakta

Pri tem načini poznamo dve vrsti sintetizatorjev to sta:

- artikularni sintetizatorji: tu je fizični model zasnovan na podlagi opisa fiziologije in akustike govornega trakta
- formantni sintetizatorji: prenosna funkcija je opisana s formantnimi frekvencami in formantnimi amplitudami - umetno rekonstrukcijo formantnih značilnosti dosežemo z vzbujanjem niza resonatorjev in antiresonatorjev (periodično vzbujenje za zveneče glasove in vzbujaje s šumom za nezveneče glasove)

Za dobro generiranje potrebujemo dober matematični model govornega trakta, pri tem pa je poraba pomnilnika majhna. Slabosti sta kompliciran matematični model pri artikularnih sintetizatorjih in težavno nastavljanje formantnih parametrov pri formantnih sintetizatorjih.

2. Z združevanjem predhodno posnetih enot

Predhodno posnete enote so vzete iz naravnega govora, kot so: alofoni, difoni, trifoni večji deli pogosteje uporabljanih besed, celih besed ali kombinacijo teh. Potrebujemo torej bazo osnovnih enot, s katerimi bomo tvorili govor. Pri manjših osnovnih enotah je poraba pomnilnika manjša, vendar pa so pri večjih enotah rezultati boljši. Zaradi spektralnih nezveznosti, ki nastanejo pri lepljenju osnovnih enot, uporabljamo postopke, ki te nezveznosti zgladijo. Najbolj razširjeni postopki, ki se uporabljajo pri obdelavi signala, so:

- večpulzno linearno napredovanje (LPC),
- obdelovanje signalov sinhrono z osnovno periodo (PSOLA in WSOLA),
- kombiniranje harmonskih in šumnih modelov (HNS, MBR-PSOLA).

Najpogosteje se uporablja algoritem TD-PSOLA (time domain PSOLA), tu združevanje osnovnih enot poteka v časovnem prostoru. Obstaja pa tudi algoritem FD-PSOLA (frequecy domain PSOLA), pri katerem poteka združevanje v frekvenčnem prostoru in daje nekoliko boljše rezultate kot TD-PSOLA, vendar je procesorsko zahtevnejši.

3 Govorec

Z razvojem in čedalje večjo uporabo računalniške tehnologije se je pokazala potreba po sistemih, ki omogočajo komunikacijo med človekom in strojem v naravnem jeziku. Za to so potrebni tako sistemi za razpoznavanje in razumevanje govora, kakor tudi sistemi za sintezo govora. Raziskave na teh področjih potekajo v svetu že preko 30 let. Razvoj je prišel tako daleč, da so takšni sistemi uporabni tudi v praksi.

V Sloveniji se trenutno ukvarjata s sintezo govora predvsem dve raziskovalni skupini: Odsek za inteligentne sisteme, ki deluje v okviru Inštituta Jožef Štefan v Ljubljani in Laboratorij za umetno zaznavanje na Fakulteti za elektrotehniko v Ljubljani.

Na Institutu Jožef Stefan je 1993. leta S. Weilguny izdelal sistem za izgovorjavo izoliranih slovenskih besed. A. Dobnikar je razvil model za napovedovanje slovenske stavčne intonacije in trajanja premorov. Leta 1995 je bil zasnovan nov sintetizator slovenskega govora, sposoben samodejnega pretvarjanja poljubnih slovenskih besedil v govor imenovan Govorec. Govorec je sintetizator govora, združljiv z Microsoftovim Speech API, kar mu omogoča enostavno integracijo v obstoječe programe za branje. Po namestitvi Govorca že nameščeni programi dobijo novo možnost branja v slovenskem jeziku, kar je za uporabnika najudobneje.

Glavni namen sistema je olajšati branje slepim in slabovidnim osebam pri delu z računalnikom. Ti uporabniki najlaže zaznavajo informacije, ki prihajajo iz računalnika, v zvočni obliki. Zaradi obilice tekstovnih informacij na internetu obstaja velika potreba po sistemu za branje besedil v slovenskem jeziku.

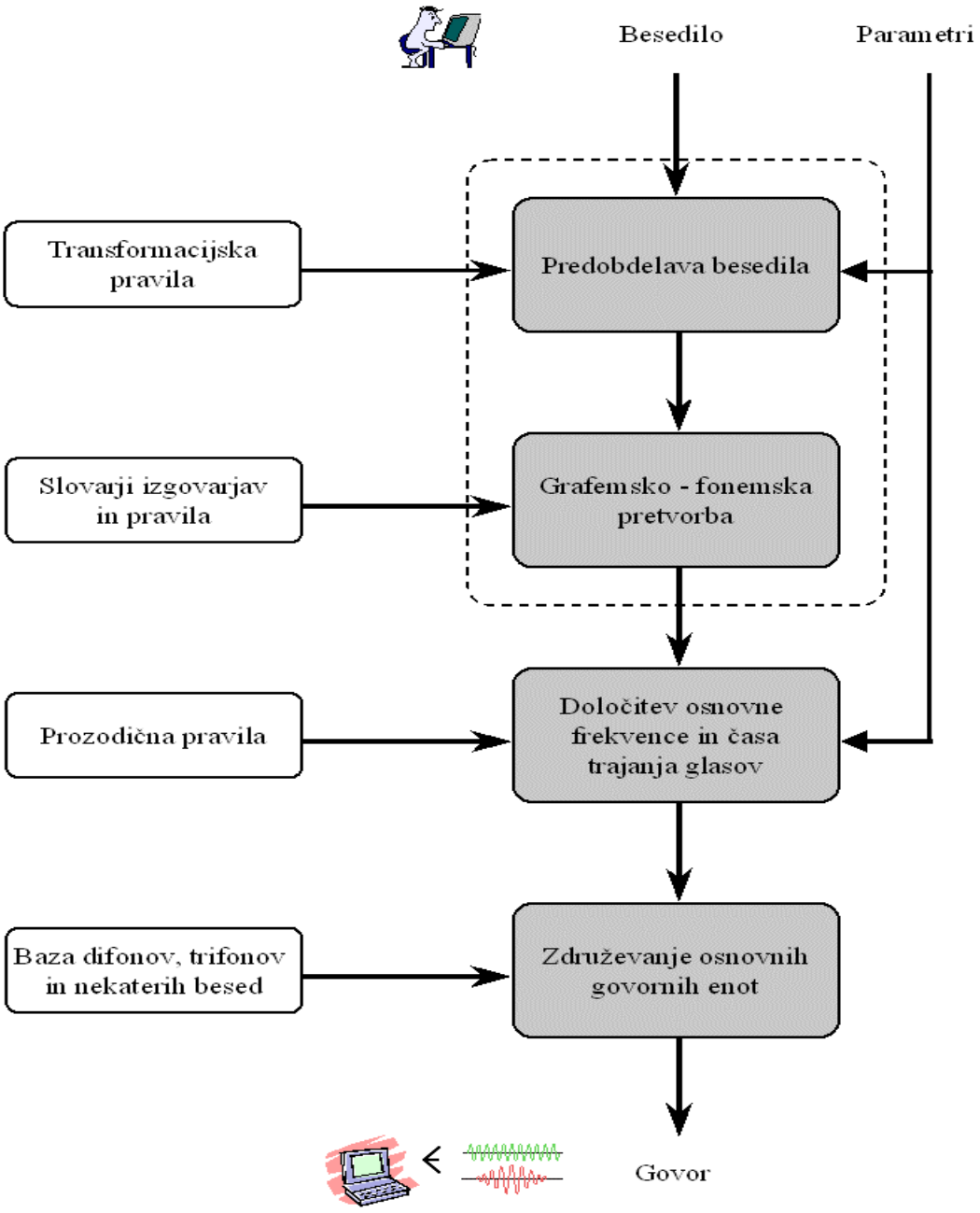
3.1 Zgradba sintetizatorja govora

Sistem je zasnovan modularno, kar omogoča enostavno popravljanje posameznih delov sistema, ki so povsem neodvisni drug od drugega.

Vhod v sintetizator je poljubno slovensko besedilo predstavljeno v digitalni obliki. V hierarhični arhitekturi sistema je na prvem mestu modul za predobdelavo besedila, sledita mu grafemsko - fonemski modul in modul za nastavljanje prozodičnih parametrov. Kot zadnji nastopa modul za združevanje osnovnih govornih enot, katerega rezultat je digitalni zapis umetnega govora, ki ga predvajamo preko zvočne kartice računalnika.

Sintetizatorju lahko nastavljamo celo vrsto parametrov]. Tako lahko spreminjamo način predobdelave besedila in sicer glede na vrsto besedila, ki ga sintetiziramo (splošna besedila, kratka sporočila, imena oken, menijev in ikon, formule, časopisni članki.). Pri nastavljanju prozodičnih

parametrov določamo hitrost govora (normalna, hitra, počasna ali katerakoli vmesna stopnja), osnovno frekvenco, mesto, število in trajanje premorov, način naglaševanja besed, način spreminjanja časa trajanja glasov, izbiramo lahko med različnimi modeli nastavljanja prozodičnih parametrov.



Slika 4 - shema sistema Govorec

3.1.1 Predobdelava besedila

V postopku predobdelave vhodnega besedila najprej pretvorimo različne formate besedila v ASCII zapis. Besedilu odstranimo nepotrebne znake, ki ne vplivajo na izgovorjavo besedila: presledki, simboli in elementi, ki v besedilo ne sodijo (HTML ukazi, kontrolne značke). Zaradi združljivosti s SAPI *Govorec* podpira tudi kontrolne značke, ki omogočajo vnos ukazov v besedilo. Z njimi lahko izboljšamo prozodične parametre. Pri vsem tem si pomagamo z najrazličnejšimi seznami pravil in slovarji, ki jih lahko po potrebi tudi dopolnjujemo

V tej fazi se pretvori tudi posebne skupine znakov v njihov grafemski zapis: razvijemo kratice in krajšave, pretvorimo številke, številke in datume v besede, posebej obravnavamo elektronske naslove, telefonske številke in matematične izraze. Pri pretvorbi poskušamo ugotoviti tudi pravo obliko (spol, sklon) razvite besede. Besedilo za tem razbijemo na stavke, stavke na besede in besede na zloge. Vsaki besedi s pomočjo naglasnih shem in pravil določimo mesto naglasa.

Na koncu predobdelave je celotno besedilo razdeljeno na stavke, stavki pa na besede. Besede tako vsebujejo le 25 grafemov slovenske abecede in informacijo o ločilu ter naglasu.

3.1.2 Grafemsko-fonemska pretvorba

Grafemsko-fonemsko pretvorbo lahko opišemo v treh korakih:

- pogledamo, ali se iskana beseda nahaja v slovarjih izgovarjav,
- če jo najdemo, prepišemo prevod besede v fonetični zapis, pri čemer dodatno upoštevamo koartikulacijo med sosednjimi besedami,
- besedam, ki jih ni v slovarjih izgovarjav, določimo mesto naglasa, temu sledi pretvorba v fonetični zapis.

Slovarje najbolj pogostih besed smo razdelili na več podslovarjev: številniki, lastna imena, akronimi, kolokacije, splošne besede (najobsežnejši del), besede iz danega področja uporabe.

Avtomatsko pretvorbo v fonetični zapis opravimo z večjim številom kontekstno odvisnih pravil. Pri nastavljanju naglasnega mesta besed si pomagamo še s seznami naglašenih in nenaglašenih enot. Glede na to, da je za slovenski jezik značilno prosto mesto naglasa, ki se ga naučimo hkrati z učenjem jezika in besed, je to vse prej kot enostavno opravilo.

Rezultat grafemsko-fonemske pretvorbe je množica stavkov, razdeljenih na besede, ki so zapisane v fonemski obliki in vsebujejo mesto ter tip naglasa.

3.1.3 Določanje prozodičnih parametrov

Ta modul vsakemu fonemu določi trajanje in višino osnovnega tona glede na naglas in mesto naglasa v besedi. Na začetku vsakemu fonemu priredimo osnovno trajanje in frekvenco, nato pa ju postopoma spreminjamo do končne vrednosti.

Zato lahko postopek nastavljanja prozodičnih parametrov razdelimo na tri korake:

- nastavljanje trajanja,
- nastavljanje osnovne frekvence,
- določitev mesta in trajanja premorov.

Modeliranje trajanja

Govornim enotam v trajanju ene besede sprva priredimo inherentne dolžine, ki jih dobimo kot vsoto inherentnih dolžin glasov, vsebovanih v besedi. Dejavniki, ki jih obravnavamo so: ime in vrsta glasu, glasovni kontekst, položaj glasu znotraj besede, vrsta zloga, naglašenost zloga, položaj zloga v besedi. Ko se besede vključujejo v večje govorne enote, se skrajšujejo ali podaljšujejo v skladu z zahtevami višjenivojskih prozodičnih pojavov. Pri tem upoštevamo različne parametre, kot so položaj besede v frazi, izbrano hitrost govora in dolžino besede, izraženo s številom zlogov v besedi.

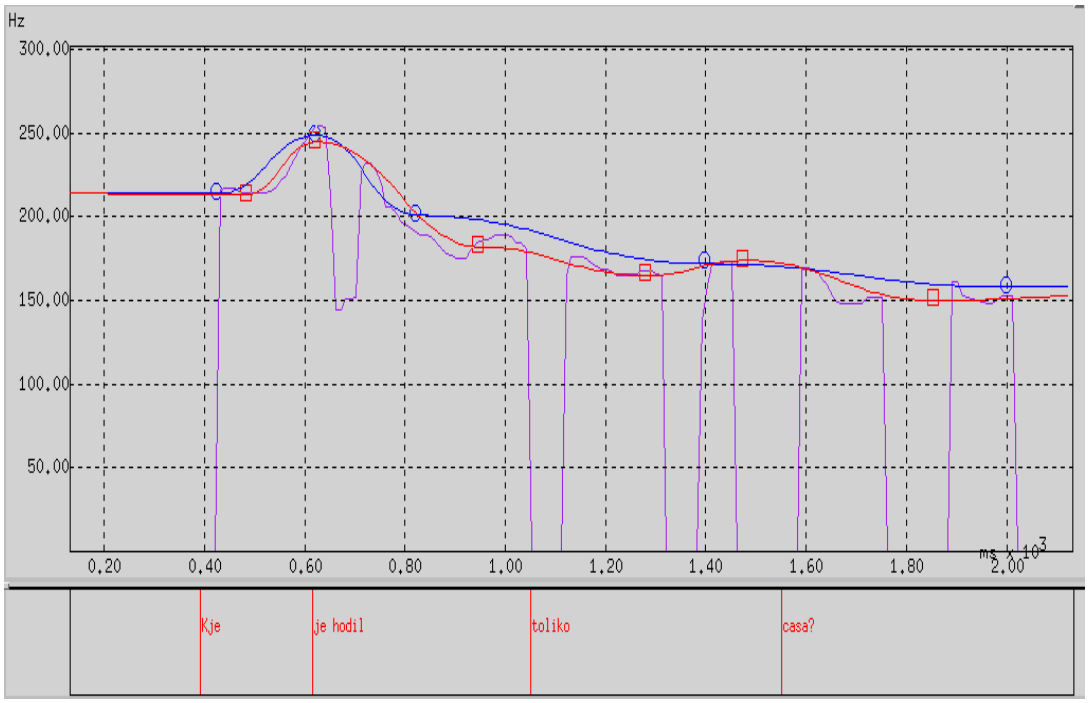
Modeliranje osnovne frekvence

Govorec uporablja superpozicijski model za določanje osnovne frekvence. Potek osnovnega tona v intonacijskem segmentu je definiran kot vsota globalne, ki je hkrati nosilna komponenta in posameznih lokalnih komponent. Globalna komponenta je določena z vrsto intonacijskega segmenta in položaja ter dolžine intonacijskega segmenta v sestavljenih oblikah povedi. Lokalna komponenta je definirana za poudarjene besede v intonacijskem segmentu.

Pri tonemskem naglaševanju ton znotraj naglašenega zloga narašča, po dosegu tonskega vrha pa začne padati v odvisnosti od vrste besede in vrste naglasa.

Najprej se lotimo problema na nivoju besede. Upoštevamo število zlogov, vrsto besede ter mesto in vrsto naglasa. Glede na relativni položaj posameznega zloga glede na mesto naglasa posebej obravnavamo prednaglasni, naglasni in ponaglasni zlog.

Stavčno intonacijo oblikujemo glede na vrsto povedi (povedna, vprašalna, vzklična). Uporabimo enega od treh značilnih potekov ovojnice osnovne frekvence.



Slika 5 - Rezultat modeliranja osnovne frekvence za stavek: "Kje je hodil toliko časa?". Kvadratki predstavljajo posnetek naravnega govora v sistemu INTSINT, krogci pa karakteristične točke umetno generiranega poteka osnovne frekvence.

Določitev mesta in trajanja premorov

Nazadnje določimo še premore v besedilu. Glede na dolžino premorov in položaj v besedilu ločimo štiri skupine:

- premori pri naslovih (so najdaljši),
- premori na koncu povedi,
- premori v povedi na mestih ločil in
- premori v povedi na mestih ritmičnih delitev (običajno pred vezniki *in*, *pa*, *ter* in sicer v daljših stavkih).

3.1.4 Združevanje difonov

Osnovne govorne enote, ki jih uporablja Govorec, so difoni. Difon je sklop dveh sosednjih fonemov. Sestavljen je iz druge polovice prvega in prve polovice drugega fonema. Difon vsebuje informacijo o glasovnem prehodu prvega fonema v drugega. Baza vsebuje 1156 difonov.
Vsak difon ima označeno sredino prehoda, na zvenečih delih pa so označene tudi periode osnovne frekvence.

Difone združujemo s postopkom TD-PSOLA (Time Domain Pitch Synchronous Overlap-Add synthesis), ki temelji na razčlenjevanju signala v zaporedje prekrivajočih se kratkotrajnih signalov. Metoda dovoljuje kvalitetne spremembe frekvence in trajanja govornega signala neposredno na časovnem signalu in je računsko poceni.

Lepljenje osnovnih period signala poteka s pomočjo Hanningovega okna, ki ima vrh na mnogokratniku osnovne periode signala. Okna so vedno daljša od osnovne periode, tako se sosednja okna prekrivajo in pri lepljenju signalov seštevajo.

Ob povezovanju prav tako upoštevamo spremembo trajanja in osnovne frekvence signala. Signal podaljšamo tako, da vrinemo zadostno število osnovnih period, če je signal periodičen, v nasprotnem primeru pa vrinemo šum (š) ali tišino (p, t, k). Frekvenco spremenimo tako, da dolžino periode ustrezno spremenimo.

Pomanjkljivost algoritma so spektralne nezveznosti na mestih lepljenja signala. Nezveznosti poskušamo odpraviti z linearno interpolacijo signala v časovnem prostoru.

3.2 Uporaba Govorca

Za delovanje potrebuje Govorec vsaj 8 Mb prostora na disku. Program deluje pod vsemi inačicami operacijskega sistema Windows od 95 naprej. Računalnik mora imeti primerno zvočno kartico (npr. Sound Blaster).

Sistem Govorec vsebuje dva programa: SiTalk in SiRead. Vsak ima svoje okence. SiTalk se uporablja za splošno delo z besedili, SiRead pa stalno teče v ozadju in ob pritisku določenih tipk prebere označeno besedilo.

3.2.1 SiTalk

SiTalk je splošen program za branje besedil. Primeren je za branje sproti vpisanega besedila. Besedilo najprej natipkamo v okencu programa, nato pa ga program prebere. Program omogoča nastavitve hitrosti branja, frekvence govornega signala in glasnosti. Z drsniki Hitrost, Višina in Glasnost nastavljamo hitrost branja, frekvenco glasu in glasnost. Pri vrednostih Hitrost in Višina pomeni manjša vrednost počasnejše branje oz. nižjo frekvenco (nižji glas).

Program omogoča uporabo poljubnega sintetizatorja govora, ki je Microsoft Speech API 4.0 združljiv. Na dnu okna v področju Sintetizator Govora (Mode) je možna izbira sintetizatorja govora. Število izbir je odvisno od števila sintetizatorjev govora nameščenih na računalniku. Privzet je slovenski sintetizator govora Govorec.

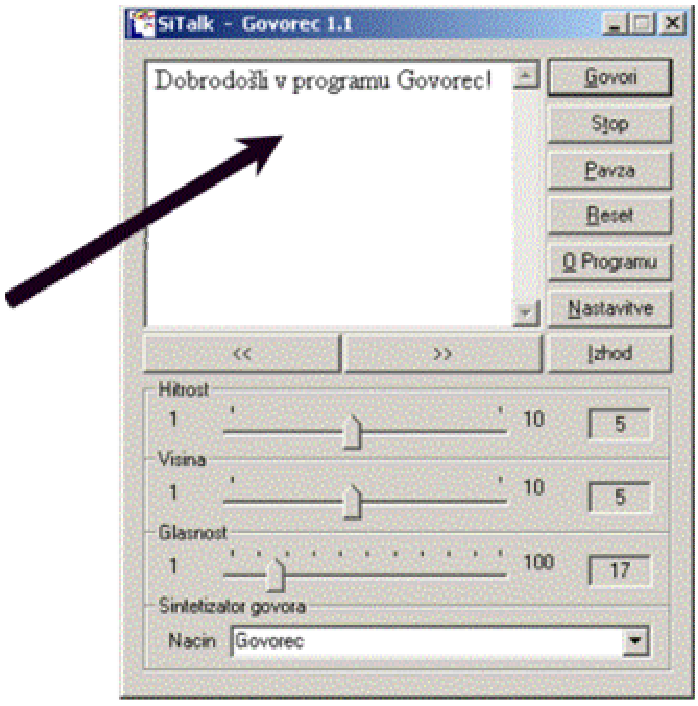
Primer uporabe programa SiTalk:

1. Zaženemo program SiTalk (sitalk.exe).

2. V prostor, označen s puščico, kliknemo z miško in natipkamo poljubno besedilo. Besedilo lahko tudi prenesemo v okence kot običajno v Windowsih.

3. Kliknemo gumb Govori ali pritisnemo kombinacijo tipk Alt+g in program SiTalk prične z branjem besedila.

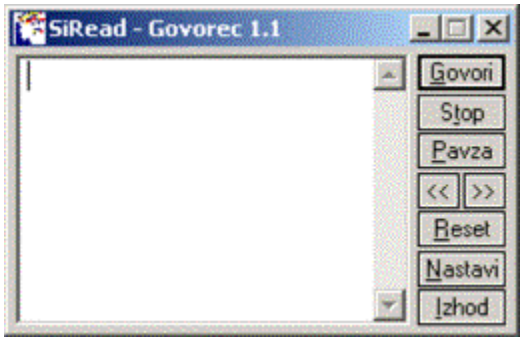
4. Med branjem uporabimo gumbe Pavza in Stop oziroma ustrezne kombinacije tipk za začasno zaustavitev ali prek



Slika 6 - SiTalk

3.2.2 SiRead

SiRead je program za branje besedila, ki se trenutno nahaja v odložišču (clipboard). Program je vedno nad vsemi ostalimi okni, pa tudi ostale funkcije so posebej prilagojene slepim in slabovidnim, tako da z nekaj tipkami začnemo brati izbrano besedilo v poljubnem programu. Za uporabo tega programa ni potrebna uporaba miške.

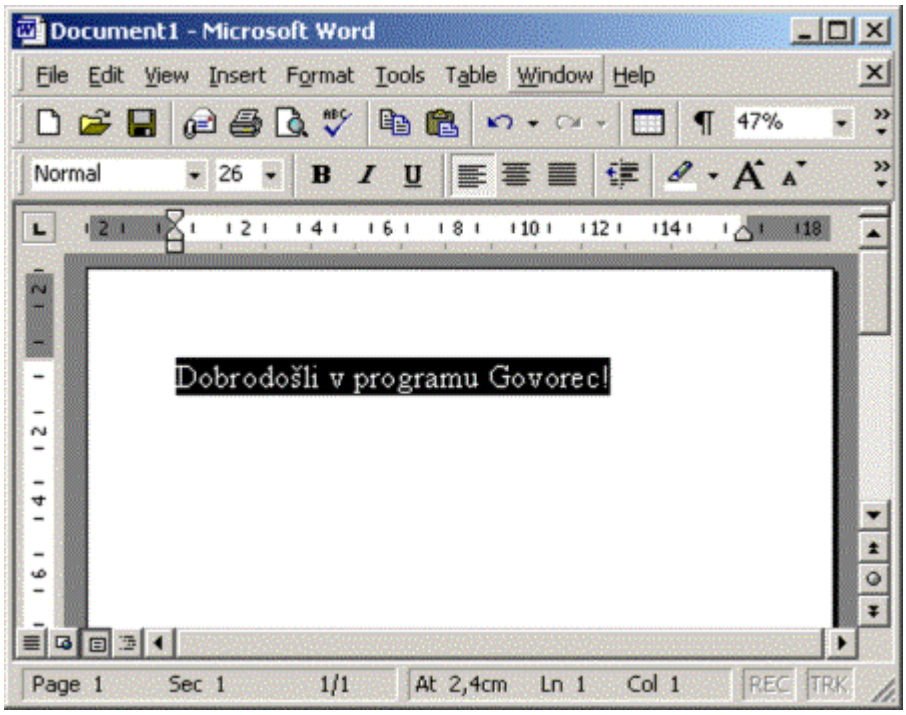


Slika 7 - SiRead

Primer uporabe programa SiRead

1. Zaženemo program SiRead (siread.exe).

2. Odpremo program za delo z besedili (WinWord, Notepad itd.). Lahko odpremo tudi program za delo z internetom (Internet Explorer, Netscape) in se postavimo na neko stran.



Slika 8 - programsko okno WinWord

3. Označimo besedilo, ki ga želimo prebrati. To lahko storimo tako, da držimo pritisnjeno tipko Shift in s smernimi tipkami označimo besedilo, lahko pa pritisnemo Ctrl + a, v tem primeru bomo označili celotno besedilo.

4. Pritisnemo Ctrl + C (Copy). S to kombinacijo tipk skopiramo označeno besedilo v odložišče.

5. Pritisnemo Ctrl + Shift + R in besedilo, ki smo ga označili, se prebere. Branje lahko prekinemo s kombinacijo tipk Ctrl + Shift + T.

4 Tuji TTS sistemi

Sinteza govora je v zadnjih letih zelo napredovala in se dodobra zelo izpopolnila. Sinteza govora ima dandanes že tako kvaliteto, da bi jo lahko primerjali z naravnim govorom. Sama kvaliteta sinteze govora verjetno ne bi bila tako visoko, če ne bi bilo na tržišču tako velika konkurenca različnih sistemov za sintezo govora. Nekateri sistemi so že tako izpopolnjeni, da nudijo izbiro kot so kakšnega spola naj bo oseba, ki nam bo tvorila glas ali pa z kakšnim naglasom naj ta oseba govori ter celo kakšne starosti naj ta oseba bo (otrok, odrasel). V najini seminarski bova predstavili najbolj uporabljene tuje sisteme za sintezo govora. To so: Neospeech, AT&T in Nuance.

4.1 Neospeech

NeoSpeech je hitro rastoča in vodilna v jezikovnih tehnologijah ter je aplikacija za mobilni, podjetni in izobraževalni trg. Nudijo visoko kvalitetni sistem za sintezo govora, ki je zmožen tvoriti govor v angleškem jezik, evropskem jezike in v večinah azijskih jezikih.

Njihov vodilni izdelek iz področja sinteze govora je VoiceText, ki tvori izjemno visoko kvalitetni človeški govor iz napisanega besedila. Trenutno ima sistem na voljo enajst različnih popolnih glasov, ki lahko tvorijo besedilo v Ameriški, Korejski, Japonski in Kitajski jezik.

Demo verzija programa se nahaja na spletni strani *http://www.neospeech.com/-demo/demo_text.php*.

Slika 9 - Neospeech demo

1. Izbereš jezik in glas v okencu "TTS Voices", izbiraš lahko med ameriško angleščino, korejščino, japonščino in kitajščino.
2. V okence vneseš tekst, ki ga želiš procesirati. Maksimalno število znakov pri demo verziji je 400 angleških znakov in 200 azijskih znakov.

3 Klik na gumb "Listen to Speech 64Kbps" pod tekstovnim oknom sintetizira avdio datoteko ASF (Advanced Streaming Format).

4. Ta datoteka se potem avtomatsko predvaja v Windows Media Player-ju.

4.2 AT&T

Na spletni strani AT&T laboratorijev *http://public.research.att.com/~ttsweb/tts/demo.php#top* je na voljo demo verzija njihovega TTS programa.

Slika 10 – AT&T demo

Demo verzija je osnovana na treh korakih. V prvem koraku nastavimo glas in jezik, v drugem koraku vnesem tekst in v zadnjem koraku sintetiziramo govor, ki ga lahko ali samo predvajamo, lahko pa tudi shranimo.

Omogoča izbiro različnih jezikov: ameriško angleščino, latinsko ameriško španščino, francoščino, nemščino, britansko angleščino. Vsak jezik pa ima na voljo tudi izbiro govorca (moški, ženska).

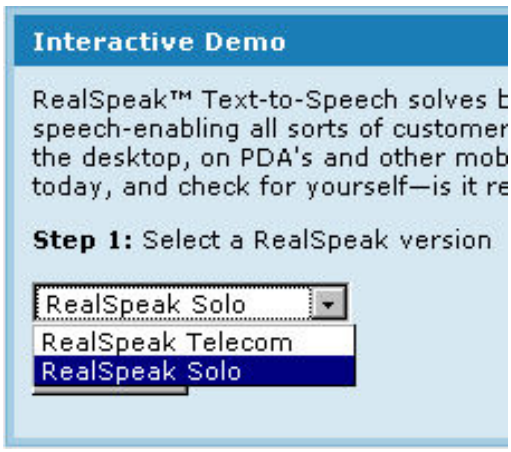
4.3 Nuance - Realspeak

RealSpeak je programska oprema, ki pretvarja tekst v visoko kakovosten moški ali ženski govor. Njegova dobra lastnost je ta, da omogoča uporabo v avtomobilskih navigacijskih sistemih, branje zaslonov za slepe in slabovidne ali pa poveča klicne usluge.

Ponujajo različne rešitve:

- **RealSpeak Telecom** je program, ki spremeni tekst v čist človeški govor v več kot 20 jezikov in 30 glasov v različnih naglasih in stilih govora. Uporablja se predvsem za govorne in mrežne aplikacije
- **RealSpeak Solo** rešitev, ki je narejen posebej za to, da poveča kakovost vgrajenih pogovornih aplikacij. Avtomobilskim, pisarniškim in mobilnim aplikacijam daje možnost govora.
- **RealSpeak Word** uporablja izjemen pristop k tekst-v-govor tehnologiji, da doseže največjo kvaliteto govora iz slovarja besed in jezikovnih posebnosti ter omogoča učencem jezika, da slišijo kako morajo biti besede izgovorjene.
- **RealSpeak Mobile** je družina tekst-v-govor izdelkov namenjenih za mobilne telefone, ki omogočajo branje SMS sporočil in elektronske pošte.

Na spletni strani podjetna Nuance *http://www.nuance.com/realspeak/demo/* je mogoče preizkusiti programa RealSpeak Telecom in RealSpeak Solo.

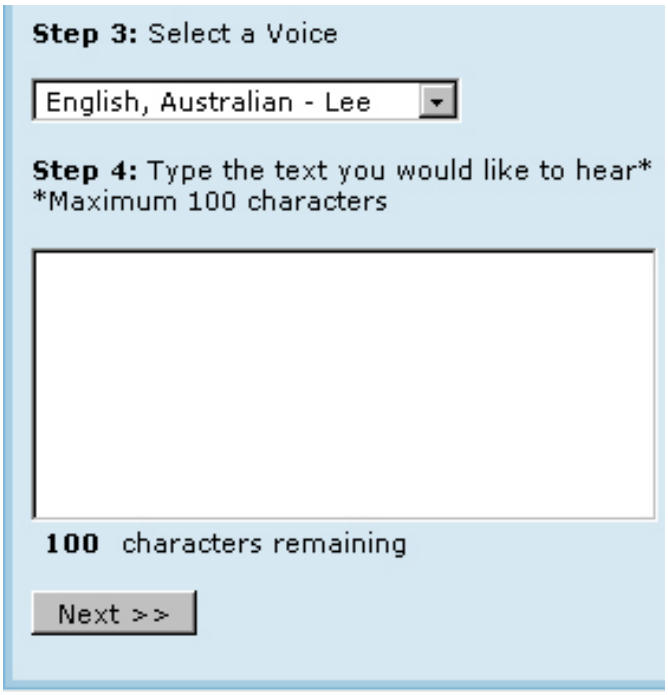


Slika 11 – 1. korak: izbira verzije ReakSpeak

V prvem koraku izberemo verzijo demo programa RealSpeak, kjer lahko izbiramo med RealSpeak Solo in RealSpeak Telecom.

Slika 12 – 2. korak: izbira jezika

V drugem koraku izberemo jezik. Izbiramo lahko med veliko jeziki; različicami angleščine, nemščino, italijanščino, francoščino, poljščino, portugalščino, španščino, celo kitajščino in japonščino.



Slika 13 - 3. korak: izbira galsu in 4. korak: vnos teksta

V tretjem koraku izberemo želeni glas; moški ali ženski. Pri nekaterih jezikih, predvsem pri različicah angleščine, nemščini in francoščini lahko izbiramo ženski ali moški glas, pri manj razširjenih jezikih pa nimamo možnosti izbire.

V četrtem koraku vnesemo v polje želeni tekst. V demo verziji je količina besedila omejena na 100 znakov.

Z klikom na gumb »next« se generira in predvaja wav datoteka.

5 Ocenjevanje kvalitete TTS sistemov

Zaradi velikega števila TTS programov, ki so na voljo danes, tako kot poskusne verzije, zastonjske aplikacije ali kot plačljivi komercialni programi, je odločitev katerega preizkusiti in nenazadnje kupiti ali uporabljati težka. V tem poglavju sva opisali nekaj testov s katerimi lahko ocenimo kakovost TTS sistemov.

Glavni faktorji, ki vplivajo na kakovost TTS sistemov so razumljivost/jasnost, naravnost in začetek v procesnem postopku. Pod jasnost spadajo lastnosti kot so količina govora, ki ga uporabnik lahko razume in kako hitro postane uporabnik utrujen kot poslušalec. Naravnost je faktor, ki poudarja pomembnost, da je produciran govor čim bolj podoben oz. enak naravnemu človeškemu govoru. Začetno procesiranje se nanaša na možnost sistema, da se inteligentno spopada z drugače človeku običajno rabo v tekstu kot nap. okrajšave, številke, homograme ipd.

Jasno je, da so ti faktorji medsebojno odvisni. Napake v začetnem procesiranju vodijo seveda do manj jasnega govora, kar pa posledično pomeni, da je govor manj naraven.

Poleg zgoraj omenjenih faktorjev seveda obstaja še cela množica drugih, ki prav tako vplivajo na kvaliteto in uporabnost TTS sistemov. Faktorji, ki so v povezavi s sistemom, so uporaba virov, enostavna uporaba vmesnika, podpora za različne platforme, uporaba dodatnih orodij, itd… Seveda pa je eden pomembnejših faktorjev tudi denar.

Tip aplikacije: različne aplikacija imajo različne zahteve glede TTS sistemov. Glas, ki bi bil primeren za sisteme, ki so namenjeni zabavi in razvedrilu, ni primeren za aplikacije v bančništvu.

Jezik in naglasi: sistemi, ki pretvarjajo tekst v govor lahko nudijo tudi različice jezikov. To je uporabno predvsem za angleški jezik, ki ima britansko, ameriško, avstralsko različico.

Namen uporabe: pomembno vprašanje je tudi namen uporabe TTS sistema. Potrebno je izbrati tak sistem, ki najbolje zadovolji potrebe in zahteve uporabnika. Tako nap., če bo sistem uporabljen za branje elektronske pošte s pomočjo telefona, potem je ena pomembnejših karakteristik, da je govor razumljiv preko telefonskega omrežja.

Osnovne tri faktorje, ki vplivajo na kakovost sintetiziranega govora in so bili predstavljeni že zgoraj, lahko preverjamo za različne sistem z različnimi testi.

5.1 Testi jasnosti/razumljivosti

Testi na osnovi jasnosti se ukvarjajo z možnostjo razpoznavanja kaj je bilo izgovorjeno oz. sintetizirano. Ti se ne ukvarjajo toliko s samo naravnostjo govora, čeprav je le ta pomembna in vpliva na razumljivost.

5.1.1 Test razumevanja:

Najpomembnejša lastnost uporabnih TTS sistemov je ali uporabniki razumejo kar so slišali. Pogosta slabost sistemov z nižjo kvaliteto je ta, da uporabniki lahko sledijo govoru, vendar zelo težko razumejo njegov pomen. Kvalitetni sistemi morajo predstaviti besedilo na tak način, da ga uporabniki z lahkoto in pravilno razumejo.

Test se izvaja na tak način, da se testirancem predvaja odlomek teksta (seveda s pomočjo TTS sistema), nato pa se jim postavlja vprašanja, ki se nanašajo na predvajani tekst. Količino pravilnih odgovorov poda okvirno oceno o razumljivosti sistema.

5.1.2 Fonetični testi

Fonetični testi se ukvarjajo s prepoznavanjem natančnih zvokov znotraj določene besede, ki je sintetizirana. Navadno se ne ukvarjajo z razumevanjem celotnih besed, fraz ali stavkov.

• DRT test (diagnostic rhyme test) je test razumevanja soglasnikov, ki se nahajajo na prvem mestu v besedi. Uporabnikom predvajajo besedne pare, ki se razlikujejo samo po prvi črki, in ti morajo potem povedati katero besedo so slišali. Nap. v angleščini dense – tense, v slovenščini miza-niza.

• MRT test (modified rhyme test) je test razumevanja zadnjih delov besed. Nap. v angleščini bath – bass.

5.1.3 Transkripcijski testi:

Transkripcijski testi se ukvarjajo s prepoznavanjem besed v stavku, ki nimam pomena. Tako se testiranci ne morejo nanašati na kontekst besedila.

• SUS test (semanticaly unpredictable sentances): uporabniki morajo prepisati slišano besedilo, ki nima pomena oziroma povezave, tako da nimajo možnosti, da bi lahko iz pomena sklepali na določeno besedo, ampak se lahko zanašajo samo na signal. Nap. angleški stavek »the thought ran towards the pink sadness« nima nobenega pomena, tako da če bi ga predvajali osebi in jo prosili da stavek ponovi, bi dobili dober občutek o tem kako razumljiv je bil govor.

5.2 Testi naravnosti

• MOS (mean opinion score) je subjektivna točkovna metoda. Uporabniki ocenjujejo kvaliteto govora za različne sisteme, običajno za sintezo istih stavkov. Običajno uporabniki ocenjujejo govor z ocenami od 1 do 5, kjer je ocena 1 zelo slabo in ocena 5 zelo naravno. Če je test natančno izveden, so lahko rezultati zelo zanesljivi. Seveda pa je MOS test zelo relativen, saj je odvisen od tega ali uporabniki poslušajo govor preko slušalk, zvočnikov ali po telefonu, kakšna je kvaliteta

zvočnikov ali slušalk, akustičnost prostora, demografski faktorji, … Ta test zahteva tudi veliko vzorčenje, če želimo pridobiti realne rezultate, vsaj 10 uporabnikov in vsaj 50 stavkov.

Vsak stavek je predvajan na vseh sistemih, vendar uporabniki ne vedo kateri sistem je bil kdaj uporabljen. Za vsak drug stavek, se mora vrstni red sistemov zamenjati, tako je manj možnosti, da uporabnik postane pristranski.

• Force-choice ranking je metoda, kjer ne ocenjujemo na podlagi primerjave. Ponavadi se sisteme ocenjuje medsebojno, ker je lažje ocenjevati primerjalno, zato se ta metoda, kjer je potrebno oceno določiti brez primerjave, manj uporaba.

5.3 Testi začetnega procesiranja

Test funkcionalnosti je najbolj objektiven test. Za ta test je smiselno uporabiti stavke, ki vsebujejo večpomenski kontekst, pri katerih pričakujemo anomalije. Primer je simbol pomišljaja »-».

Angleški primer:

> My social security number is 345-23-9803.
> My social security number is three four five twothree nine eight zero three.

V tem primeru so številke govorjene brez posebnih pavz, številke naštevamo.

> From 1975-1979, I was studying at theuniversity.
> From nineteen seventy five to nineteen seventynine, I was studying at the university.

V tem primeru se pomišljaj pretvori v besedo »to«.

> Our telephone number is (607) 266-7025.
> Out telephone number is six owe seven, two sixsix, seven owe two five.

V tem primeru se pomišljaj pretvori v premor.

V ocenjevanje TTS sistemov je smotrno vključiti sposobnost večpomenskosti-konteksta, ker lahko v nasprotnem primeru izberemo sistem, ki ne glede na primernost pretvori pomišljaj v besedo »to« (v angleščini) oz. v besedo »do«. Cilj normalizacije teksta je zadeti pravilno ravnovesje med tem da stvari izpeljemo pravilno in tem, da ne posplošujemo preveč.

5.4 Preizkus programov na osnovi testov

V tem delu poglavja bova opisali, kako so se odrezali različni sistemi za sintezo govora glede na osnovne tri faktorje, ki vplivajo na kakovost sintetiziranega govora. Ta preizkus sva izvedli tako, da sva v izbranih sistemih za sintezo govora vpisale določeno besedilo in potem ugotavljale v katerih je bila tvorba besedila v govor boljša iz vidika jakosti/razumljivosti, naravnosti in začetnega procesiranja.

5.4.1 NeoSpeech

Tabela 1 - Ocenjevanje kakovosti sistema NeoSpeech

Napisano besedilo	Tvorjeni govor
I had absolutely nothing to do last week – it was wonderful.	Pri tem besedilo se je sistem odrezal odlično, saj je besedilo zelo razumljivo, poslušalec razume celotno govorjeno besedilo. Pri tvorbi govora so ustrezno upoštevana ločila kot premori. Samo tvorjenje govora deluje precej naravno in je le malenkost zaznati, da gre za sintezo govora.
"I'll never give in to you," he murmured.	Pri tem primeru smo ugotovili, da pri sintezi govora pride, do rahlih težav, ko uporabljamo več ločil v bližini črk, saj jih sistem poskuša povezati z besedilom. Zato dobimo zelo nejasen začetek govora. Če pa je samo eno ločilo pa je ta pravilno uporabljen. Tudi ta sistem ne deluje popolnoma naravno.
David Elkins arrived in San Francisco Wednesday afternoon to address striking janitors who walked out last Tuesday highlighting the plight of low-salary earners who are living in one of the most expensive cities in the US.	Ta sistem je besedilo, ki je napisano kot celota, brez ločil tvoril v govor tako, da je bil govor zelo hiter, brez ustreznih premorov. Zaradi tega je zelo težko slediti celotnemu besedilu, čeprav je izgovorjava besed odlična. Zaradi vseh teh vzrokov je sinteza govora zelo nenaravna.
I know the butcher, the baker and the candlestick maker.	Tukaj je izgovorjava zelo jasna. Govor deluje zelo naravno, ker so upoštevana pravilno ločila kot premori.
Could you pass me the sugar, please?	Izgovorjava pri tem primeru je jasna, besede so izgovorjene razločno in razumljivo. Sinteza govora deluje zelo naravno, saj gre za upoštevanje vprašalnega stavka, prav tako vejice kot premor.
Two dozen eggs 4 quarts of Milk Orange Juice Bread (Whole Grain)	V tem primeru gre za zelo nejasen govor, saj sistem ne zazna da gre za naštevanje ampak poveže besedilo v celoto. Tako dobimo zelo hitro naštevanje brez premorov, ker ni ločil. Edini premor, ki ga zaznamo je pri oklepajih. Prav zaradi tega deluje ta sistem tu zelo nenaravno.
My social security number is 345-23-9803.	Tovorjeni govor ni zadovoljiv, ker se v primeru telefonske številke pomišljaj izgovori kot »dash«
From 1975-1979, I was studying at theuniversity.	Tovorjeni govor je zadovoljiv, ker pomišljaj sistem pretvori v besedico »to«
Our telephone number is (607) 266-7025.	Tovorjeni govor je zadovoljiv, ker pomišljaj sistem pretvori kot presledek/pavzo

5.4.2 AT&T

Tabela 2 - Ocenjevanje kakovosti sistema AT&T

Napisano besedilo	Tvorjeni govor
I had absolutely nothing to do last week – it was wonderful.	Ta sistem v tem primeru je sicer zelo jasen in razumljiv, vendar deluje zelo nenaravno, ker so ločila upoštevana z zelo majhnim intervalnim premorom in je govor zelo hiter.
"I'll never give in to you," he murmured.	Govor je nejasen na začetku, ko želi povezati med seboj ločila in besede. Drugače je govor zelo jasen, ker upošteva ustrezne premore in naglase. Zaradi slabega začetka govora ta deluje rahlo nenaravno.
David Elkins arrived in San Francisco Wednesday afternoon to address striking janitors who walked out last Tuesday highlighting the plight of low-salary earners who are living in one of the most expensive cities in the US.	Govor je zelo jasen, Tudi če je zelo dolg in brez ločil, sistem sam smiselno umešča premore in smiselno razdeli celotno besedilo v govor, ki deluje dokaj naravno.
I know the butcher, the baker and the candlestick maker.	Govor je jasen, ker ima pravilno naglaševanje in povezovanje besed. Deluje kar se da naravno.
Could you pass me the sugar, please?	Govor je jasen vendar deluje zelo sintetično, besedilo med ločili pove kot dve samostojni celoti tako ne deluje povsem kot vprašanje. Zaradi tega tudi ne deluje naravno.
Two dozen eggs 4 quarts of Milk Orange Juice Bread (Whole Grain)	Govor je manj jasno, saj besedilo pove zelo hitro in jo zazna kot celoto in ne kot naštevanje. Zato deluje zelo nenaravno, poleg tega pa besedilo v oklepaju pove ločeno od ostalega.
My social security number is 345-23-9803.	Tovorjeni govor ni zadovoljiv, ker se v primeru telefonske številke pomišljaj izgovori kot »dash«
From 1975-1979, I was studying at the university.	Tovorjeni govor je zadovoljiv, ker pomišljaj sistem pretvori v besedico »to«
Our telephone number is (607) 266-7025.	Tovorjeni govor je zadovoljiv, ker pomišljaj sistem pretvori kot presledek/pavzo

5.4.3 Nuance

Tabela 3 - Ocenjevanje kakovosti sistema Nuance

Napisano besedilo	Tvorjeni govor
I had absolutely nothing to do last week – it was wonderful.	Govor je bil razumljiv. Pri tvorbi je imel rahle vmesne presledke. Govor ne deluje ravno naravno, vendar pa upošteva ustrezne premore.
"I'll never give in to you," he murmured.	Govor je zelo jasen, ker je izgovorjava tekoča. Poleg tega upošteva ločila in nima težav tudi, če so ta ločila blizu skupaj. Vendar ne deluje naravno, ker se sliši veliko vmesnih preskokov med besedami.
David Elkins arrived in San Francisco Wednesday afternoon to address striking janitors who walked out last Tuesday highlighting the plight of low-salary earners who are living in one of the most expensive cities in the US.	Govor zelo jasen in razumljiv. Kljub dolgemu besedilu ni prehiter in se slišijo ustrezni premori, zato je tudi govor bolj naraven. Vendar ni čist, ker je slišanih vmesnih preskokov med besedami.
I know the butcher, the baker and the candlestick maker.	Govor je jasen. Upoštevajo se ustrezni premori pri ločilih. Tudi v tem primeru govor ni čist, ker ima vmesne preskoke med besedami.
Could you pass me the sugar, please?	Govor je jasen. Dobro se razločijo posamezne besede. Vendar ne zveni povsem naravno kot da je vprašanje.
Two dozen eggs 4 quarts of Milk Orange Juice Bread (Whole Grain)	Govor je manj jasno, saj besedilo pove zelo hitro in jo zazna kot celoto in ne kot naštevanje. Poleg tega pa ne upošteva oklepaja pri izgovorjavi. Zato deluje zelo nenaravno.
My social security number is 345-23-9803.	Tovorjeni govor je zadovoljiv, ker se v primeru telefonske številke pomišljaj pretvori v pavzo
From 1975-1979, I was studying at theuniversity.	Tovorjeni govor je zadovoljiv, ker pomišljaj sistem pretvori v besedico »to«
Our telephone number is (607) 266-7025.	Tovorjeni govor je zadovoljiv, ker pomišljaj sistem pretvori kot presledek/pavzo, poleg tega število v oklepaju prepozna kot »area code«

5.4.4 Primerjava različnih TTS sistemov

5.4.4.1 Tuji TTS sistemi

Testi za različne TTS sisteme so pokazali, da se v večini primerih najbolje odreže sistem NeoSpeech, saj tvori govor, ki je zelo jasen in poslušalec razume celotno govorjeno besedilo. Pri tvorbi govora so ustrezno upoštevana ločila kot premori. Samo tvorjenje govora deluje precej naravno. Razen v primerih, ko imamo dolgo besedilo brez ločil ter ločila, ki so pisana blizu skupaj. Slabše pa se je prav tako odrezal pri zaznavanju pomišljaja kot delitev številk, kjer ga pretvori v besedico »dash«.

Sledi mu sistem Nuance, pri katerem je tvorba govora prav tako razumljiva. Vendar pa lahko zaznamo rahle vmesne preskoke med besedami, zato govor ne deluje ravno naravno. Vendar upošteva ustrezne premore pri ločilih in tudi kljub dolgemu besedilu ni prehiter. Tudi nima večjih težav pri zaznavanju števil in ločili, ki povezujejo števila.

Najslabše od preizkušenih sistemov je odrezal AT&T sistem, kjer je tvorjenje govora sicer zelo jasno in razumljivo, vendar pa deluje zelo nenaravno, ker so ločila upoštevana z zelo majhnim intervalnim premorom in je govor zelo hiter. Nima pa težav pri zaznavanju daljših besedil, razen v primeru ko gre za besedilo, ki predstavlja seznam in pa ne zazna kdaj gre za trdilni stavek in kdaj za vprašalni stavek. Prav tako pride do slabšega prepoznavanja ločil pri številkah, kjer prav tako kot sistem NeoSpeech pomišljaj tvori kot besedico »dash«.

5.4.4.2 Govorec

Slovenski sistem za pretvorbo teksta v govor je v primerjavi s preizkušenimi tujimi aplikacijami na precej nižji stopnji. V smislu razumljivosti, je sistem še dokaj dober, medtem ko na ravni naravnosti razočara. Govor je zelo nenaraven in poslušalca kmalu utrudi. Uporabnik se močno skoncentrira na samo razločevanje posamezne besede, da ne razmišlja o samem pomenu govorjenega besedila. Naglaševanje besed velikokrat ni korektno. Sistem ne upošteva intonacije na podlagi ločil (kadar gre za vprašanje, se glas na koncu stavka ne zviša, pri vzkliku glas ni višji in močnejši, vejice na razume kot premor). Simbol »-« (vezaj) prebere ne glede na primer kot minus.

Sam program ne omogoča izbire različnih glasov, lahko pa nastavljamo hitrost, višino in glasnost govora.

6 Zaključek

Hiter razvoj računalništva in informatike je posegel na številna področja človeškega udejstvovanja. Tudi obdelava naravnih jezikov se temu ni mogla izogniti. Tako smo dobili številna orodja, brez katerih si danes težko zamišljamo tudi povsem običajna opravila.

Napredke je mogoče zaznati tako na področju sinteze govora, razpoznavanja in strojnega prevajanja. Vse te tehnologije so ključne v jezikovnih tehnologijah in so v velikem porastu, saj si današnja družba ne more več predstavljati življenja brez njih. To pa zato ker omogočajo hitrejšo komunikacijo, izobraževanje, združevanje narodov in rešitev nekaterih ovir, kot je omogočanje slepim in slabovidnim za delo z računalnikom ter lažje prebiranje vsebin.

Vendar kljub temu pa, da so tehnologije že zelo napredovala in so bili narejeni že veliki dosežki, pa je lahko trditi, da je še veliko neraziskanega. Zato veliko podjetij težijo k razvoju novih in vse bolj naprednih tehnologij, da bi zagotovila svojim uporabnikom orodja in okolja, ki bi zadovoljila njihovim potrebam. Zato lahko pričakujemo še bolj revolucionarne razvoje na tem področju tako v tujini kot pri nas.

7 Literatura

[1] http://laps.fri.uni-lj.si/dps_arhiv/seminarske/indihar/sinteza.htm

[2] http://ai.ijs.si/govorec/

[3] http://ai.ijs.si/govorec/solomon/slide-2.html

[4] http://ai.ijs.si/govorec/govorec-navodila.html

[5] http://www.cs.cmu.edu/~jure/pubs/govorec-is01.pdf

[6] http://en.wikipedia.org/wiki/TTS

[7] http://public.research.att.com/~ttsweb/tts/demo.php

[8] http://www.neospeech.com/demo/demo_text.php

[9] http://www.nuance.com/realspeak/demo/default.asp

[10] http://www.bell-labs.com/project/tts/

[11] http://www.tmaa.com/tts/ Evaluating%20TTS%20Systems%20White%20Paper%2010-02.pdf

[12] http:// www.tmaa.com/tts/ SpeechWorks%20White%20Paper%20Writing%**20for**%20TTS.pdf

[13] http://www.acoustics.hut.fi/~slemmett/dippa/thesis.pdf

[14] http://public.research.att.com/~ttsweb/tts/faq.php#TechWho

[15] ROMIH, M. Jezikovne tehnologije. *Življenje in tehnika*, 2003, let. 54, št. 1, str. 20-29