

Univerza v Ljubljani  
Naravoslovnotehniška fakulteta  
Oddelek za tekstilstvo  
Grafična tehnologija

Seminarska naloga

# **STROJNO PREVAJANJE GOVOR-GOVOR**

Predmet: Jezikovne tehnologije

Avtorici: Polona Kuzman  
Nina Šturm

Ljubljana, januar 2006

# KAZALO

1. Uvod.....	3
2. Predstavitev projekta.....	3
3. Prevajalnik .....	4
4. Govorne tehnologije.....	5
4.1 Avtomatska razpoznavna govora.....	5
4.2 Prevajanje.....	8
4.3 Avtomatska sinteza govora.....	10
4.3.1 Postopki avtomatske sinteze govora .....	10
4.3.2 Sinteza govora pri projektu.....	11
5. Zaključek.....	15
6. Literatura in viri .....	16

# 1.UVOD

Ko se na sestankih in kongresih srečuje veliko ljudi, ki govorijo različne jezike, so za uspešno sporazumevanje nujno potrebni prevajalci, ali pa uporaba enega splošnega jezika, kot je na primer angleščina. Vendar pa so razmere postale skoraj kritične v današnji Evropi. Od maja 2004 je v Evropski uniji 25 držav in 20 uradnih jezikov. Prej 110 jezikovnih parov za prevajanje je naraslo na 380 parov in stroški prevajanja vrtoglavo rastejo. Presegli so 5 milijard evrov na leto.

Raziskave so pokazale, da prebivalci Evropske unije želijo informacije v svojem jeziku: spletna strani, časopisi, formularji, poročila in obvestila ... Raziskave so pokazale tudi, da število spletnih strani z neangleško vsebino narašča.

Zato je v sklopu šestega okvira razvojnega programa EU nastal projekt TC-STAR, ki vključuje univerze in strokovnjake ter podjetja z dolgoletnimi izkušnjami. Povezuje jezike kot je naglasna angleščina, evropska španščina in mandarinščina. Trije jeziki, ki so pomembni za evropski trg. Trajanje projekta je tri leta in je trenutno v zadnji fazi<sup>1</sup>.

Ta seminar opisuje zasnovo najinega projekta, ki se zgleduje po projektu TC-STAR, vendar pa bi bil prirejen za slovenski jezik.

## 2. PREDSTAVITEV PROJEKTA

### **Osnovna ideja najinega projekta je**

Zasnova aplikacije, ki bi omogočala prevajanje angleškega jezika v slovenski na osnovi govora. Zato bi bilo potrebno razviti vmesnik, ki bi omogočal razpoznavo govora v angleškem jeziku, vnos prevedel s strojnim prevajnikom, ter rezultat v slovenskem jeziku podal v obliki govora, torej sintetiziral govor.

### **Razlogi:**

- Strojno prevajanje na ravni človeka – govor je naravna lastnost človeka.
- Tehnologija danes to omogoča.

### **Prednosti takšnega sistema:**

- Uporabniku ni potrebno znanje uporabe računalnika – prijazen in logičen grafični vmesnik
- Omogoča proste roke, ni potrebno tipkanje ali branje.
- Uporabljajo ga lahko tudi slepi, slabovidni in hendikepirani.
- Pomoč pri učenju sloveščine za tujce.
- Če imamo aplikacijo na dlančniku, jo imamo vedno s seboj.
- Nadgradnja v interpretacijski telefon – prevajanje jezika klicočega v jezik klicanega. Uporabo takšnih storitev napovedujejo za konec tega desetletja.

### 3. PREVAJALNIK

Prevajalnik deluje na osebnih računalnikih in dlančnikih. Upravljamo ga z govorom, prepoznava ukaze v slovenskem in angleškem jeziku, kot na primer:

*Odpri prevajalnik / Start translator* (zažene se aplikacija in prikaže se grafični vmesnik).

*Prevedi / Translate* (aktivira se proces razpoznavanja govora za angleški jezik)

Pojavi se ikona mikrofona in sistem je pripravljen za vnos.

Izgovorimo stavek, ki ga želimo prevesti.

*Stop* (pove programu, da je vnos zaključen, ikona mikrofona izgine)

Na zaslonu se izpiše izgovorjeni stavek.

Ukaz *Poprava / Cancel* omogoča popravo, če stavek ni pravilno prepoznan, ponovno izgovorimo stavek, lahko pa popravimo stavek tudi z ročnim vpisom s pomočjo tipkovnice.

*Nadaljuj / Proceed* ukaz posreduje stavek prevajalnemu procesu, ki stavek prevede v slovenščino.

Na zaslonu se izpiše stavek v slovenščini, sintetizator govora pa sintetizira govor, ki ga slišimo skozi zvočnike.

Ukaz *ponovno povej / say again* ponovi sintetiziran stavek, kar se lahko ponavlja.

*Začni znova / Start over* nas vrne na začetek.



Slika 1: Grafični prikaz prevajalnika

#### Sestava aplikacije:

- **grafični vmesnik s sistemom za govorno upravljanje aplikacije**

(kar omogoča uporabniku uporabo sistema, govorno upravljanje je mogoče v slovenskem ali angleškem jeziku)

- **sistem za razpoznavo govora**

(deluje za angleški jezik, razpozna akustični signal govorca in ga izpiše)

- **sistem za strojno prevajanje**

(prevaja stavke iz angleškega v slovenski jezik, vsebuje slovnična pravila za stavkotvorje)

- **sinteza govora**

(preveden stavek v slovenščini sintetizira v govor)

- **podatkovne baze in slovarji za angleški in slovenski jezik**

(baze so potrebne za razpoznavo govora, posebej so baze in slovarji za prevajanje, in posebej baze, potrebne za generiranje slovenskega jezika)

**- program, ki vse to povezuje**

(aktivira aplikacijo, prenaša vmesne rezultate od enega do drugega procesa)

## 4. GOVORNE TEHNOLOGIJE <sup>2</sup>

Tehnologija	Uporaba
Avtomatsko razpoznavanje govorca	Uporaba identifikacijskih značilnosti govorca kot gesla (poslovanje z banko na daljavo)
Kodiranje govora	Kodiranje in dekodiranje govora za njegovo učinkovito shranjevanje (govorna pošta)
Avtomatska sinteza govora	Sinteza govora iz besedila (prebiranje sporočil, teksta)
Razpoznavanje izoliranih besed odvisnega govorca	Razpoznavanje majhnega nabora besed za določenega govorca (govorno izbiranje telefonske številke)
Razpoznavanje izoliranih besed neodvisnega govorca	Razpoznavanje majhnega števila besed neodvisnega govorca (poslovanje z banko na daljavo, upravljanje računalniških aplikacij in sistema)
Razpoznavanje tekočega govora neodvisnega govorca iz omejenega slovarja	Razpoznavanje črkovanih imen, številskih nizov (preverjanje kreditnih kartic)
Razumevanje govora	Semantična analiza tekočega spontanega dialoga med človekom in strojem (avtomatsko posredovanje informacija različnih informacijskih sistemov) <sup>2</sup>

### 4.1 AVTOMATSKA RAZPOZNAVA GOVORA

**Dejstva:**

- Človek nima problemov z razpoznavanjem moškega, ženskega glasu, razume počasi in hitro izgovorjene stavke in celo mrmrajoč govor.
- Človek ni občutljiv na šum iz okolja

**Stroj je pri tem omejen, glede na te omejitve lahko delimo različne sisteme<sup>2</sup>:**

- sistem za enega ali več govorcev
- način govora (izolirane besede, tekoč govor)
- velikost slovarja, ki ga sistem lahko prepozna

**Glede na način razpoznavanja govora delimo sisteme v tri skupine<sup>2</sup>:**

- **Razpoznavanje izoliranih besed**
  - o presledek mora biti vsaj 200ms,
  - o uporaba na veliko področjih zaradi preprostosti in velike zanesljivosti,
  - o sistemi so lahko odvisni ali neodvisni od govorca
  - o slovar velikosti približno 1000 besed
  - o uporaba omejena na aplikacije, kjer beseda že predstavlja ukaz
- **Razpoznavanje vezanega govora**
  - o Vhodni signal je tekoč govor
  - o Uporaba modelov, ki jih sestavljajo izolirane besede in primerjanje celotne vhodne izgovorjave s tako zgrajenimi modeli
  - o S pazljivo izgovorjavo lahko dozežemo dobre rezultate
  - o Primer uporabe je vnos niza števil (št. tekočega računa ... )
  - o Uporaba je mogoča tam, kjer lahko uporabnika seznanimo z omejitvami sistema
- **Razpoznavanje tekočega govora**
  - o Najbolj zapleten, nobenih omejitev za uporabnika
  - o Sistem mora znati procesirati neznane časovne meje v govornem signalu
  - o Obvladati mora vse koartikulacijske pojave in slabo izgovorjavo.

#### **Slovar sistema**

Sistemi lahko prepoznajo le omejeno število različnih besed, ki sestavljajo slovar sistema razpoznavanja govora<sup>2</sup>.

- **Sistemi z majhnim slovarjem (do 99 besed)**
  - o Aplikacije za kreditne kartice, razpoznavanje telefonske številke
- **Sistemi s srednje velikim slovarjem (100 – 999 besed)**
  - o Največkrat eksperimentalni laboratorijski sistemi za razvoj sistemov razpoznavanja tekočega govora
- **Sistemi z velikim slovarjem (1000 besed in več)**
  - o Na tržišču za uporabo na osebnih računalnikih
  - o Običajno nekaj deset tisoč besed

#### **Problemi<sup>2</sup>:**

Šum lahko zniža uspešnost razpoznavanja govora tudi za več kot 40%!

Moduli za zmanjševanje šuma

- analiza in filtriranje frekvenčnega spektra
- stacionarni filtri
- adaptivno filtriranje
- preslikave s Fourierjevimi transformacijami

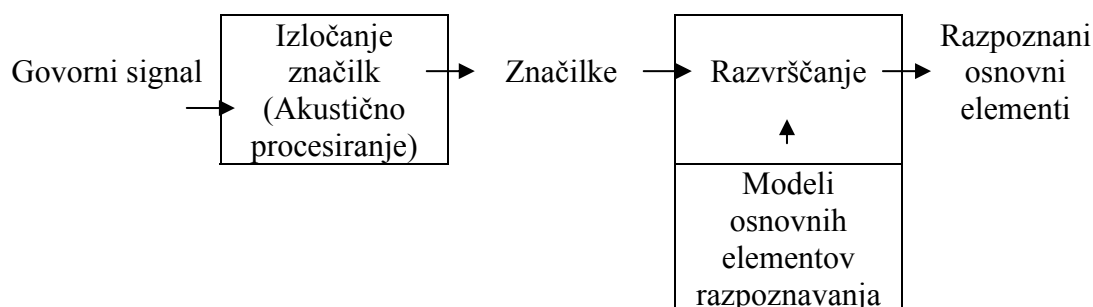
## Osnovna zgradba sistema avtomatskega razpoznavanja govora

Velika večina sistemov je zasnovanih na teoriji razpoznavanja vzorcev, kjer moramo vhodni vzorec prepoznati s pomočjo merjenja razdalj med njim in v sistemu shranjenimi poznanimi vzorci. Vzorec je lahko beseda, zlog, fonem.

V prvem koraku izločimo iz signala značilke, ki razlikujejo različne vhodne vzorce med seboj. V modulu razvrščanja značilk z merjenjem razdalj med vhodnim vzorcem in primerjalnimi vzorci (modeli osnovnih elementov razpoznavanja) določimo vzorec. Primerjalne vzorce določimo v fazi učenja sistema, kjer njihove vrednosti običajno določimo s pomočjo statističnih podatkov, dobljenih iz velike množice govornih vzorcev, ki so ponavadi zbrani v različnih bazah izgovorjav. Te so ponavadi posnete na zgoščenkah<sup>2</sup>.

Razvrščanje lahko izvedemo največkrat s pomočjo statistične odločitvene teorije, ki temelji na določitvi največje podobnosti vzorcev. Tako se izmed primerjalnih vzorcev odločimo za tistega, ki smo mu pri primerjanju z vhodnim vzorcem pripisali največjo verjetnost<sup>2</sup>.

Splošna shema<sup>2</sup>



Primerjalni vzorci so torej lahko fonemi, zlogi ali besede. Modele zlogov ali besed uporabljamo najpogosteje pri sistemih izoliranih besed. Pri prepoznavanju tekočega govora pa običajno uporabimo modele fonemov ali modele še manjših elementov govora. Na višjih nivojih procesiranja pa kot primerjalne vzorce uporabimo slovar izgovorjav in jezikovni **model**. Razvrščevalnik je tako sestavljen iz dveh modulov. V prvem določamo verjetnosti fonemov, ki predstavljajo osnovne gradnike pri kasnejšem razpoznavanju besed in stavkov. V drugem modelu pa določamo najbolj verjetno verigo besed, to je nabor besed, ki imajo največjo verjetnost, da so bile tudi dejansko izgovorjene. Sistemi avtomatskega razpoznavanja govora morajo vsebovati tudi modul za zmanjševanje šuma, prilagajanje na vhodni kanal in zmanjševanje odmeva<sup>2</sup>.

Sistem avtomatskega razpoznavanja govora običajno vključuje dve fazi: fazo učenja in fazo razpoznavanja. V fazi učenja določimo vse potrebne parametre govora in primerjalne vrednosti za vsak nivo procesiranja (npr. Primerjalne vektorje, vektorske kodne knjige, vrednosti matrik verjetnosti prehodov prikritih modelov Markova, verjetnosti jezikovnih modelov itd.) v fazi razpoznavanja pa značilke vhodnega signala primerjamo s primerjalnimi vrednostmi<sup>2</sup>.

#### 4.1.1 PROGRAMSKA OPREMA ZA AVTOMATSKO RAZPOZNAVANJE GOVORA<sup>3</sup>

- Za angleški jezik je na trgu ogromno izbire
- Dragon Naturally Speaking in drugi
- Omogočajo avtomatsko razpoznavanje govora in jih je mogoče integrirati v druge aplikacije s pomočjo: ActiveX, C++, Visual Basic
- Kompatibilni so s skoraj vsemi urejevalniki teksta.
- Program je potrebno "uglasiti" na uporabnika.
- Uporabnik se mora naučiti uporabe.

#### 4.2 PREVAJANJE

Po razpoznavanju govora in izpisu le tega sledi strojno prevajanje. V projekt sva vključil prevajalni sistem Presis, saj je zaradi upoštevanja slovničnih pravil slovenskega in drugih jezikov je Presis učinkovitejši od klasičnih prevajalskih orodij, še posebej v kombinaciji z elektronskimi slovarji, ki jih je razvilo podjetje Amebis<sup>4</sup>.

Prevajalni sistem Presis ponavadi ne prevaja povsem pravilno. Prevod je namenjen približnemu razumevanju originalnega besedila ali osnovi za nadaljnje prevajanje. Vendar je kakovost prevoda odvisna tudi od dejavnikov, na katere lahko vplivamo. Za boljše prevajanje se moramo držati navodil<sup>4</sup>:

- JavaScript in piškotki (Cookies) sta nujna za pravilno delovanje teh strani.
- Stavke in lastna imena moramo pisati z veliko začetnico (imena jezikov, dni v tednu in mesecev v slovenščini niso lastna imena).
- Besedilo za prevajanje morajo biti slovnično pravilna, saj to pomaga pri boljši analizi.
- Predstavitvena verzija prevajalnega sistema Presis je omejena na prevajanje besedil z dolžino največ 200 znakov. V predstavitveni verziji ni mogoče dodajanje lastnih prevodov.



Po začetni fazi, ko je bilo prevajanje večinoma na nivoju besed z nekaj osnovne stavčne analize, je Presis zdaj prešel na uporabo glagolskih predlog pri stavčni analizi in generacij in prevodov.

Predloga opiše, kako se določen glagol uporablja, pove, s katerimi predmeti se veže, katere omejitve so pri tem, katera prislovna določila so lahko ob glagolu, katere predložne zveze so tipično povezane z glagolom in podobno. Predloge so potem povezane v pomen, pri čemer so pripisane tudi potrebne stilne oznake, ki jih prevajalnik uporablja za boljše izbiranje prevodov. Da bi se izognili prevelikemu ponavljanju pri predlogah, je del informacij, ki so značilne za veliko število glagolov v posameznem jeziku, rešen programsko na nivoju prevajalnika (analizatorja oz. generatorja) in so v teh primerih v predlogah poudarjene le izjeme. Tak primer je npr. vezava s prislovnimi določili, kjer so tipično dovoljena dodatno k predlogu poljubna prislovna določila, razen prislovnega določila kraja, po katerem se vprašamo kam. Analizator in generator tudi skrbita za pravilen vrstni red prislovnih določil v posameznem jeziku.

S pomočjo predlog lahko razrešimo pri prevajanju dvournost, ki se slovnično razlikujejo, recimo glede prehodnosti/neprehodnosti, povratnosti pri slovenskih glagolih, frazne glagole v angleščini,... Prav tako je možno stalne fraze v celoti vpisati v predloge, s čimer lahko dosežemo pravilne prevode fraz. Ta zapis je veliko bolj prilagodljiv, kot bi bil zapis z enostavnim zaporedjem besed, saj so elementi predloge lahko v poljubnem vrstnem redu (znotraj pravil posameznega jezika).

Tako lahko Presis zdaj brez težav prevede tudi stavek: »Bobu je včeraj Janez rekel bob.« v »Janez called a spade a spade yesterday.«

Pri tem sicer najde tudi druge možne prevode (brez uporabe fraz), vendar da na prvo mesto rešitev, kjer predloga pokrije čim večji del stavka (postopek za razvrščanje najdenih analiz uporablja sicer več kriterijev za razvrstitev, vendar je ta med pomembnejšimi in ponavadi prevlada).

Predloge določijo tudi v stavčne člene, tako da je zaradi tega mogoče povezovati pomene glagolov s pomeni samostalnikov glede na vlogo, v kateri se pojavljajo. Tako je mogoče določiti, da sta glagol (oz. glagolski pomen) »skuhati« in samostalni »kosilo« povezana tako, da je kosilo tipično predmet v tožilniku (oz. roditeljskem v zanikanih stavkih, za kar poskrbi sam analizator in tega v vzorcih ni treba posebej pisati) pri tem glagolu, ne pa npr. osebek ali predmet v dajalniku.

Pri predlogah je možno napisati tudi tipične predložne zveze, s katerimi se povezujejo. S tem lahko po eni strani izboljšamo prepoznavanje predlog, po drugi strani pa tudi dosežemo, da se predlog ob določenem glagolu prevaja drugače, kot bi se sicer običajno.

»V našega Janeza se je zaljubila lepa Micka.« tako Presis prevede v »Beautiful Micka fell in love with our Janez.«

Presis zna predloge tudi prilagoditi različnim časom in trpniku. Morebitne izjeme (predloge, ki ne morejo uporabljati v trpniku ali se lahko uporabljajo le v trpniku, omejitve pri časih) so označene z uporabo lestvic pri predlogah. Če stavka, ki ji bil v originalu v trpniku, v ciljnem jeziku ni mogoče zapisati na ta način (to se običajno zgodi pri prevajanju iz angleščine v slovenščino, kadar je znan tudi osebek, ki je bil v angleščini naveden z »by«), ga generator pretvori v tvorni način. Primer za to je prevod iz: »John was made happy by Mary.« v »Mary je osrečila Johna.«

Posebnost slovenščine so tudi glagoli, ki se dobijo v neosebni rabi drug pomen. Tak primer je: »Gre za ta problem.«, kar Presis s pomočjo predloge prevede v »It is this

problem.« oz. kot druga možnost »It is about this problem.« Najde pa seveda tudi kopico drugih možnosti, recimo »It goes behind this problem.«

Predloge so lahko take, da kot del zahtevajo nov glagol. Tipičen primer so predloge za modalne glagole (oz. v slovenščini tudi modalne prislove, recimo »lahko« in »rad«). V takem primeru zna Presis sestaviti predlogi. Primer za tak prevod je: »Why could John fall in love with Mary?«, kar prevede v »Zakaj bi se John lahko zaljubil v Mary?«

Ocenjujejo, da bo za zadovoljivo delovanje prevajalnika potrebnih med 10.000 in 20.000 vnesenih predlog na jezik, jezik pa bo večinoma pokrit pri okoli 100.000 predlogah. Težava pri ugibanju predlog so v slovenščini povratni glagoli in glagoli v neosebni rabi, v angleščini pa frazi glagoli, zato imajo ti pri vnosu predlog prednost (skupaj z najbolj pogostimi glagoli)<sup>5</sup>.

## 4.3 AVTOMATSKA SINTEZA GOVORA

Gre za posredovanje informacij, ki jih s pomočjo signala stroj posreduje uporabniku – človeku<sup>2</sup>.

Prednosti uporabe tovrstne sinteze pri komunikaciji med človekom in strojem so<sup>2</sup>:

- Uporabnik zlahka razume govornega sporočila, ki mu jih posreduje stroj brez predznanja in zahtevanja večje koncentracije, saj je govor najpogostejše in najučinkovitejše sredstvo za sporazumevanje med ljudmi.
- Uporabnik lahko sprejme informacijo, tudi ko opravlja druge aktivnosti npr. hodi, dela na drugih objektih...
- Posredovano sporočilo ne potrebuje medija za svoj zapis papirja ali zaslona, kar je lahko pri posredovanju daljših sporočil tudi slabost.

Prvi sintetizator, ki je zmožni generirati tekoč govor se je imenoval VODER, izdelal ga je H.W.Dudley (Bell Laboratories) leta 1939. Sistem se je krmilil s pedalom in desetimi tipkami, s pomočjo katerih je lahko operater spreminjal osnovno harmonsko frekvenco in karakteristike pasovnoprepustnih filtrov<sup>6,7</sup>.

### 4.3.1 POSTOPKI AVTOMATSKE SINTEZE GOVORA

Postopke avtomatske sinteze danes razdelimo v tri skupine<sup>2</sup>:

1. Sinteza s pomočjo **kodiranja govornega signala**. Za tvorjenje govornih sporočil uporabljamo vnaprej posnet govorni signal, ki je lahko zaradi zmanjšanja pomnilniških zahtev kodiran. Kvaliteta sestavljenega sporočila je odvisna od ujemanja značilk govora med posameznimi sestavljenimi besedami ali frazami. Med akustične značilke prištevamo ovojnico, amplitudo, osnovno frekvenco in hitrost izgovora. Če shranimo daljše fraze, stavke dosežemo večjo naravnost in razumljivost sintetiziranega govora. Če pa posnamemo manjše enote, kot so zlogi in fonemi, lahko tvorimo večje število različnih besed in stavkov, vendar je kvaliteta slabša.

V praksi se shranjujejo besede in fraze, nato se pri generiranju sporočila uporabi ustrezne besede v kombinaciji s frazami. Pri tovrstni sintezi pa je

največji problem nenaravnost govora, saj se izgovorjen stavek precej razlikuje od zaporedja besed, ki so izgovorjena izolirano. To rešujemo tako, da posamezno besedo posnamemo z različnimi poteki osnovne harmonske frekvence (naraščajoča, konstantna in padajoča). Vendar sintetiziran govor še vedno ni popolnoma zadovoljiv, saj je potrebno rešiti tudi problem naglasa v stavku, ritem in intonacijo, ki so odvisni od sintaktičnih in semantičnih dejavnikov.

Problem, ki se še pojavi pri tovrstni sintezi je razširitev besed. Če želimo dodati novo besedo, to ne moremo storiti brez dodatnega poprejšnjega snemanja. Torej moramo imeti zmeraj na voljo ustrezen prostor z enakim akustičnim okoljem in referenčnega govorca, ki bo izgovoril novi material. Če starega govorca nimamo na voljo, mora novi govorec izgovoriti, celotni slovar, čeprav smo želeli dodati le eno besedo.

2. Sinteza zasnovana na postopku **analiza-sinteza** je izboljšana različica prejšnje, saj obsega manjše število podatkov, ki so potrebni za sintezo govora. Tu se prvotno lotimo analize izgovorjave besed, posameznih fraz in določimo zaporedje značilk, ki podajajo časovno spreminjanje značilnosti analiziranega govora. Pri sintezi nato sestavljamo značilke posameznih osnovnih enot - fonemov,, besed, fraz. S pomočjo sestavljenih zaporedij značilk sintetizator generira govorni signal. Na ta način tvorimo želeno govorno sporočilo. Ker moramo za posamezno osnovno enoto shraniti le zaporedja značilk, je obseg podatkov, ki jih potrebujemo za sintezo, mnogo manjši kot pri postopku kodiranja govora. Prednost je tudi ta, da je enostavneje spreminjati značilnosti sintetiziranega govora (višina tona, hitrost govora), saj lahko to dosežemo s spreminjanjem vrednosti parametrov (vektorjev značilk). Kot zaporedja značilk lahko uporabimo LPC in PARCOR koeficiente, ki jih določimo s pomočjo linearne predikcije. Sisteme, ki izvajajo postopke določanja parametrov govora (vektorjev značilk) imenujemo vokoderji in z njihovo pomočjo lahko v veliki meri zmanjšamo pomnilniške kapacitete.
3. Sinteza s **pomočjo pravil**. Tvorimo govorni signal s pomočjo fonetičnih in lingvističnih pravil.

#### 4.3.2 SINTEZA GOVORA PRI PROJEKTU

Namen sinteze pri najinem projektu je spremeniti napisan tekst v zvočni signal, ki ga zaznavamo kot govor, torej narediti sistem, ki to omogoča (TTS sistem – Text-To-Speech system).

Besedilo je potrebno pretvoriti v fonetični zapis in nato določiti prozodično informacijo iz samega besedila. Obeh korakov ne moremo popolnoma ločiti med seboj, saj sta naglas in način izgovora tesno povezana. Veliko vpliv na naglas in izgovor pa ima semantika – pomen izgovorjenega. Tako je na primer za pravilno razumevanje besedne zveze gori na gori gori, zelo pomemben naglas in način izgovora. Takšno znanje je težko vključiti v sistem avtomatske sinteze govora<sup>2</sup>.



Slika 2: Shema sistema pretvorbe besedila v govor

- Normalizator pregleda besedilo, določi ločila, izpiše okrajšave in denarne zneske npr. 124.50 SIT v sto štiriindvajset tolarjev in petdeset stotinov oziroma številke 234 v dvesto štiriintrideset. Izhod normalizatorja besedila ima obliko zaporedja črk in ločil.
- Naloga sintaktično/prozodičnega analizatorja je razčlenitev vhodnega zaporedja črk in ločil tako, da je posameznim delom besedila moč smiselno prirediti intonacijo in ritem. Izhod sintaktično/prozodičnega analizatorja sestavljata tekst in sintaktično/prozodične oznake..
- Oba predstavljata vhod v modul izgovorjave, ki tvori fonetično transkripcijo vsake besede za vhodni tekst. Fonetično transkripcijo določimo s pravili za pretvorbo črk v foneme. Pri tem moramo uporabiti še slovar izjem, ki vsebuje tiste besede in njihove fonetične transkripcije, ki se dovolj pogosto pojavljajo in za katere bi uporaba pravil dala napačno fonetično transkripcijo. Tako modul najprej preveri, če je vhodna beseda zapisana v slovarju izjem. Če ni, tvori s pomočjo pravil ustrezno fonetično transkripcijo. Največje probleme pri tvorjenju fonetične transkripcije predstavljajo tujke in tuja imena. Za pravilno transkripcijo le-teh bi bilo potrebno definirati novo množico pravil, ki pa lahko nasprotujejo že definiranim pravilom.
- Izhod modula izgovorjave je zaporedje fonemov s pridruženimi sintaktično/prozodičnimi oznakami. Pretvorbo zaporedja fonemov v sintetiziran govor lahko izvedemo s fonemskim sintetizatorjem.

Na tržišču obstaja kar nekaj sistemov pretvorbe besedila v govor, predvsem za angleški jezik. Nekateri od teh sistemov omogočajo tudi izbiro moškega, ženskega in

otroškega glasu. Sistemi običajno uporabljajo foneme ali difone in izvajajo sintezo s pomočjo formatnega ali linearno predikcijskega sintetizatorja<sup>2</sup>.

- AT&T
- ScanSoft (Lernout & Hauspie)
- Lucent Technologies
- Microsoft Reader
- Elan Informatique
- Read Please<sup>8</sup>

Tudi na našem tržišču se lahko pohvalimo s tovrstnim sistemom za pretvorbo besedila v govor, ki se imenuje **Govorec** in je nastal v okviru Skupine za inteligentne sisteme na Institutu Jožef Štefan. Glavni namen sistema je olajšati branje slepim in slabovidnim osebam pri delu z računalnikom. Za delovanje potrebujete 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)<sup>6</sup>.

### **Delovanje Govorca**

Sistem sestavljajo trije neodvisni moduli: modul za grafemsko fonemsko pretvorbo, modul za nastavitev prozoičnih parametrov in modul za združevanja osnovnih govornih enot (difonov). Vhod v sintetizator predstavlja poljubno besedilo. Besedilo se sprva na podlagi pravil in fonemskega slovarja, ki vsebuje 500.000 besed, pretvori v ustrezen fonemski zapis. Sledi določanje osnovne frekvence in trajanja za posamezne foneme, nato se oblikuje še celotna stavčna intonacija. Sam postopek sinteze govornega signala temelji na lepljenju osnovnih govornih enot s pomočjo algoritma TD-PSOLA, ki zagotavlja visoko naravnost sintetiziranega govornega signala. Za osnovne govorne enote so izbrali difone. Govorni signal se nato predvaja na zvočni kartici (SoundBlaster), oziroma se generira datoteka v WAV formatu. Sistem je tudi Microsoft Speech API 4.0 kompatibilen, kar omogoča uporabo Govorca v katerikoli aplikaciji v Oknih, če je le-ta Speech API združljiva. Dosedanje meritve kažejo, da je sistem razumljiv nad 95%, vendar delež še naraste, ko se poslušalec navadi na glas sintetizatorja<sup>6</sup>.

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<sup>6</sup>.

- 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.

- SiRead je program za branje besedila. 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.

Govorec se lahko uporablja z ostalimi namenskimi programi za slepe in slabovidne, kot sta programa **Narrator** in **Jaws**<sup>6</sup>.

- Narrator je sestavni del operacijskega sistema Windows 2000. Ob zagonu bere naslove tekočih oken in informacijo na okviru oken. Program se nahaja na Start->Programs->Accessories->Accessibility->Narrator. Zaženemo program Narrator. Pritisnemo na gumb Voice... in v novo odprtem oknu Voice Settings izberemo Govorec, SLOVENIAN, Jožef Stefan Institute. Vrednosti Speed, Volume in Pitch nastavimo na 5 in pritisnemo gumb OK. Narrator začne brati slovensko.

#### IZBOLJŠAVE GOVORCA

- nova baza.
- boljša pravila (ucenje na slovarju).
- boljša prozodija.

## 5. ZAKLJUČEK

Za izdelavo sistema strojnega prevajanja govor v govor za slovenski jezik bi bilo potrebno integrirati sistem za razpoznavanje govora, ki ga ne bi bilo potrebno umerjati na uporabnika, izboljšati obstoječi sistem za strojno prevajanje v slovenski jezik (vključiti več pravil slovenskega jezika) in prenoviti Govorca (nova baza, boljša pravila, boljša prozodija).

## 6. LITERATURA IN VIRI

1. PROJEKT TC-STAR: <<http://www.tc-star.org/>>, 15.12.2005
2. Kačič, Z.: *Komunikacija človek – stroj*, Maribor : Fakulteta za elektrotehniko, računalništvo in informatiko, Inštitut za elektrotehniko, 1995; ISBN 86-435-0125-5
3. SAY I CAN SPEECH RECOGNITION: <<http://www.sayican.com/>>, 15.12.2005
4. SISTEM PRESIS: <<http://presis.amebis.si/prevajanje/>> 26.12.2005
5. Debeljak, M.: *Strojno prevajanje*, seminarska naloga, Ljubljana, 2005
6. SISTEM GOVOREC: <<http://ai.ijs.si/govorec/>>, 5.1.2006
7. VODER: <[http://www.obsolete.com/120\\_years/machines/vocoder/](http://www.obsolete.com/120_years/machines/vocoder/)>, 6.1.2006
8. READ PLEASE: <<http://www.readplease.com/>>, 5.1.2006
9. WIKIPEDIA: <[www.wikipedia.org](http://www.wikipedia.org)>, 15.12.2005
10. The Festival Speech Synthesis System:  
<<http://www.cstr.ed.ac.uk/projects/festival/>>, 15.12.2005
11. The MBROLA Project: <<http://www.tcts.fpms.ac.be/synthesis/>>, 15.12.2005
12. ATRC: <<http://www.utoronto.ca/atrc/reference/tech/voicerecog.html>>, 15.12.2005