

Univerza v Ljubljani
Filozofska fakulteta
Oddelek za slovenistiko

NEJC ROBIDA
Sinteza govora in Govorec 3
Diplomsko delo

Ljubljana, september 2013

Univerza v Ljubljani
Filozofska fakulteta
Oddelek za slovenistiko

NEJC ROBIDA
Sinteza govora in Govorec 3
Diplomsko delo

Mentor: doc. dr. Hotimir Tivadar

Ljubljana, september 2013

ZAHVALA

Za pomoč pri pripravi diplomske naloge se zahvaljujem mentorju doc. dr. Hotimirju Tivadarju, ki je z veseljem sprejel temo in mi jo pomagal razviti, sodelavcem na Amebisu, predvsem Simonu Rozmanu za prijetno delovno izkušnjo, svetovanje pri tehničnem delu naloge in zaupanje v moje jezikoslovne rešitve pri Govorcu 3, in svoji družini, ki me je podpirala in vodila do konca študija.

IZVLEČEK

V diplomski nalogi *Sinteza govora in Govorec 3* je povzeta zgodovina razvoja sintetizatorja govora, predstavljene so različne vrste sintez, posebej pa je izpostavljen jezikoslovni vidik razvoja sintetizatorja. Podrobneje so opisane izkušnje z delom pri Amebisovem Govorcu 3 in težave, ki so se pojavljale pri transkribiranju in urejanju slovarja izgovorjav, podana je tudi analiza nekaterih posebnih akustičnofonetičnih pojavov v govorni zbirki.

Ključne besede: sinteza govora, sintetizator, govorna zbirka, transkripcija, Govorec 3

ABSTRACT

The diploma *Speech synthesis and the speech synthesizer Govorec 3* summarizes the history of speech synthesizer development. Furthermore, different types of syntheses are presented, and particularly highlighted is the linguistic view of synthesizer development. Experiences when working with the speech synthesizer Govorec 3 from Amebis and the difficulties that have emerged while making phonetic transcriptions and editing the pronunciation lexicon are described in detail. Finally, several analyses of the distinct acoustic-phonetic phenomenon in a speech corpus are presented.

Key words: speech synthesis, speech synthesizer, speech corpus, phonetic transcription, Govorec 3

Kazalo

1 Uvod.....	1
2 Tvorjenje glasov in govorjeni jezik	2
3 Zgodovina sintetizatorjev govora.....	5
4 Sintetizator govora	7
5 Sinteza.....	9
5.2 Sintetizator govora Govorec	13
5.2.1 Sintetizatorja Govorca 1 in 2	13
5.2.3 Sintetizator Govorec 3	16
5.2.4 Govorne zbirke.....	17
5.2.5 Slovenske govorne zbirke	18
6 Vhodno besedilo	20
6.1 Slovar izgovarjav kot del Govorca.....	20
6.1.2 Slovar izgovarjav in Govorec 3	22
6. 2 Slovnična analiza pri Govorcih 2 in 3.....	23
6.2.1 Slovnična analiza in transkripcije pri Govorcu 3.....	24
6.3 Grafemska pretvorba.....	27
6.3.2 Grafemska pretvorba in Govorca 2 in 3.....	27
6.4 Transkripcija	28
6.4.1 Transkribiranje in Govorec 3	30
7 Anketa in rezultati.....	41
8 Zaključek.....	43
9 Povzetek.....	45
10 Viri in literatura.....	47
11 Priloge	49
11.1 Anketa o kakovosti in uporabnosti Govorca 2.....	49

11.1.1 Besedilo 1.....	50
11.1.2 Besedilo 2.....	50
11.2 Zgoščanka s posnetki sintetiziranega govora.....	51

Kazalo slik

Slika 1: Shematski prikaz postopka učenja in sinteze govora z uporabo prikritih modelov Markova (Vesnicer idr. 2004)	10
Slika 2: Dekompozicija glasu <i>u</i> na harmonski in šumni del (Rozman 2005: 10)	11
Slika 3: Dekompozicija glasu <i>z</i> na harmonski in šumni del (Rozman 2005: 11)	12
Slika 4: Shema sinteze besede <i>lipa</i> po postopku TD-PSOLA	14
Slika 5: Prelivanje difonov	15
Slika 6: Naglas in analiza besede <i>zavóhala</i>	15
Slika 7: Slovar izgovarjav (delovno okno)	20
Slika 8: Slovar izgovarjav (primer dvojnic)	22
Slika 9: Delovno okolje v Transcriberju	31
Slika 10: Analiza besede <i>cianid</i> (bralka)	33
Slika 11: Analiza besede <i>socialnoekonomski</i> (bralka)	34
Slika 12: Analiza besede <i>cianid</i> (bralec)	34
Slika 13: Analiza besede <i>socialnoekonomski</i> (bralec)	35
Slika 14: Analiza zvenečega <i>y</i> (bralka)	37
Slika 15: Analiza zvenečega <i>y</i> (bralka)	38
Slika 16: Analiza zvenečega <i>y</i> (bralec)	38
Slika 17: Analiza zvenečega <i>y</i> s primerom nezvenečega <i>x</i> (bralec)	39
Slika 18: Analiza primera [sl"a:bix tr"u:Sp]	40

Kazalo tabel

Tabela 1: Seznam glasov.....	5
Tabela 2: Primeri napačnih slovničnih analiz	25
Tabela 3: Primerjava izgovora členka <i>že</i> pri bralcih.....	32
Tabela 4: Primeri izgovora <i>i</i> pred samoglasnikom	35
Tabela 5: Primeri besed, v katerih si v verigi sledita dva soglasnika	36
Tabela 6: Časovna analiza zlitih glasov (bralka)	36
Tabela 7: Časovna analiza zlitih glasov (bralec)	37
Tabela 8: Številčni rezultati ankete	41

1 Uvod

Človeški govor je najpomembnejši člen medsebojnega sporazumevanja, zato je prizadevanje za strojno posnemanje glasu v času vsesplošne informatizacije pomembna znanstvenoraziskovalna dejavnost, ki je že in bo vedno bolj uporabljana tudi v tržne namene. Pri ustvarjanju sintetizatorja govora se srečamo z različnimi znanstvenimi področji v jezikoslovju, računalništvu in elektrotehniki, kar je pristen primer interdisciplinarnosti. „Sinteza govora je proces umetnega ustvarjanja govora, ki ga lahko uporabljamo za najrazličnejše namene“ (Žganec Gros 2001: 21), lahko je pripomoček slepim, samodejni odzivnik v različnih telekomunikacijskih sistemih, interaktivni sogovornik ali pa samo strojni bralec različnih besedil.

V diplomskem delu smo na kratko opisali zgodovino sintetizatorjev in pregledali razvoj slovenskih strojnih bralcev, še posebej smo se posvetili Govorcu podjetja Amebis, d. o. o., in primerjali delovanje starega Govorca z novim, ki je v fazi dokončevanja. Predstavili smo osnove slovenske fonetike, delovanje govoril in lastnosti slovenskega govorjenega jezika. Opisali smo težave, ki so se pojavljale pri izpopolnjevanju Govorca. Na koncu smo analizirali še pojavnost in lastnosti zvonečega γ. Posebna pozornost je bila posvečena predvsem praktičnim težavam pri ustvarjanju fonetičnih zapisov, saj sta tematika in problematika sintetiziranja govora preobsežni za diplomsko delo.

Pri delu z Govorcem smo se veliko ukvarjali tudi z notacijami, to so enotne in standardizirane računalniške abecede, ki s posebnimi simboli zapisujejo glasove. V diplomskem delu smo za fonetične zapise uporabljali notacijo MRPA, to smo uporabljali tudi pri transkribiranju govorne zbirke za Govorca 3. Te zapise nakazujejo oglati oklepaji ([]), ti pa se žal pojavljajo tudi v citatih, kjer smo jih zamenjali s poševnicami (/ /). Današnjo verzijo Govorca, ki je na spletu na Amebisovi strani, bomo imenovali Govorec 2 in novo Govorec 3, saj so nekateri postopki in izzivi pri ustvarjanju sinteze precej različni. Govorec 1 pa je bil last Instituta „Jožef Stefan“. Zaradi jasnosti besedo govorec uporabljamo samo za računalniški program, osebe, ki so bile posnete pri ustvarjanju govornih zbirk, pa se imenujejo bralci, saj so vsa besedila, ki so uporabljena za sintetizator govora, prebrana in niso nikoli prosto

govorjena. Diplomskega dela nismo izrecno delili na teoretični in praktični del, vse težave in ugotovitve pri delu z novim Govorcem smo navajali sproti pri vsakem poglavju.

2 Tvorjenje glasov in govorjeni jezik

Za uspešen in realističen sintetizator govora je pomembno dobro poznavanje akustične in izgovorne fonetike ter delovanja govoril – ti osnovni procesi so opisani že v Toporišičevi slovnici (Toporišič 2004) in jih tudi na kratko povzemamo. Organi, ki sodelujejo pri tvorjenju glasov, so: trebušna prepona, rebra z mišicami prsnega koša, pljuča, sapnici s sapnikom, grlo s pokrovko, žrelna stena, ustna in nosna votlina, jezik, zobje in ustnice. Za tvorjenje sta pomembna predvsem izdišni zrak in nihanje glasilk, ki z enakomernim tresenjem tvorijo zven oziroma osnovni ton (F_0), tega bomo iskali tudi pri zvenečem γ . Glasilki se treseta pri vseh zvenečih glasovih, pri nezvenečih ($p, f, t, s, c, \check{c}, \check{s}, k, h$) pa ne. Vsi zašepetani glasovi so nezveneči. Ob razprtih glasilkah in odprti nadglasilčni poti slišimo na primer nezveneči grlni x , ko se pripreta in zanihata, pa nastane zveneči grlni γ . Za raziskovanje glasov so pomembni predvsem sonogrami, iz katerih so vidni tudi formanti vsakega glasu. Formanti so frekvenčne komponente govora, s katerimi lahko izgovorimo in razločimo različne glasove. O njih so pisali že Lehiste, Petek, Toporišič, Ozbič in Tivadar.¹ Najkrepkejši so pri samoglasnikih, srednja odprtostna stopnja je pri zvočnikih in najmanjša pri nezvočnikih. Nezvočnike delimo na zapornike (p, t, k, b, d, g), zlitnike ($c, \check{c}, d\check{z}$) in pripornike ($f, s, \check{s}, h, z, \check{z}$). Glasove delimo še na netrajne (zlitniki in zaporniki) in trajne (Toporišič 2004: 41–45). Fonetično-fonološke zakonitosti so pri sami sintezi zelo pomembna osnova. Ob tem imejmo „v mislih, da je število možnih glasov neskončno ter da je odločitev fonetika, kdaj bo dele glasovne verige prepoznaval kot različne glasove in kdaj kot ponovitve istega glasu, arbitrarna“ (Golden 1996: 131). Zato je pomembno, da se s transkripcijami govornih zbirk, namenjenimi sintetizatorju, ki morajo biti zaradi poznejše samodejne programske analize enotne, ukvarja čim manj raziskovalcev hkrati, najbolje eden.

¹ O poglobljenih analizah formantov slovenskih samoglasnikov več tudi v članku Petra Jurgeca *Formant frequencies of standard slovene vowels* (slov. *Formantne frekvence samoglasnikov standardne slovenščine*). Tam svoje raziskave primerja s prejšnjimi in poda v prilogah tudi številčne rezultate. Več o odčitavanju formantov piše Jurgec še v članku *Natančnost odčitavanja formantov pri digitalnem spektrografiranju na podlagi LPC-analize*. Hotimir Tivadar piše o formantih v prispevku z naslovom *Normativni vidik slovenščine v 3. tisočletju – knjižna slovenščina med realnostjo in idealnostjo*.

Cilj same sinteze je, da je umetno tvorjen govor čim bolj podoben naravnemu. „Za človeški govor je značilno, da je nadvse spremenljiv. Načini izgovarjanja iste besede si niso nikoli povsem enaki“ (Žganec Gros 2000: 8). „Pri moških je osnovna frekvenca od 60 do 230 Hz, pri ženskah od 180 do 400 Hz /.../ Tako kot hipno ne moremo spremeniti frekvence glasilk, ne moremo hipno preoblikovati govorne cevi. Če želimo sintetizirati sintetizirani govor čim bolj naravno, moramo zagotoviti, da so zvezne spremembe tudi v sintetiziranem govoru“ (Rozman 2005: 2). Medtem ko je sinteza govora sama že zelo dodelana, ostaja največja težava „počlovečenje“ in s tem „poslabšanje“ idealnega signala do te mere, da se sliši naravno fluenten in človeški. Pri naravnosti govora se ukvarjamo tudi z lastnostmi govorjenega jezika in njegovim razmerjem do pisnega jezika; pri delu z Govorcem (1, 2 in 3) se s temi vprašanji redno srečujemo. Govorna zbirka, s katero smo se ukvarjali, temelji na istih besedilih, ki sta jih bralca brala z različnimi izgovori in naglaševanji besed. Vsaka izgovorjena beseda pa je zaradi samega sistema delovanja Govorca oblikoslovno in izgovorno določena tudi v slovarju izgovarjav, zato smo morali poleg pravilnega fonetičnega zapisa posnetkov skrbeti tudi za pravilni zapis izgovora v slovarju.² Tu smo se večinoma zanašali na pravopis, saj je „v slovenskem prostoru pogosto sinonim za reševanje kodifikacijskih reševanj v celoti. Pravzaprav je kot kodifikacijski priročnik obravnaval še pravilen izgovor in tudi vsa slovnično težja mesta slovenskega jezika“ (Tivadar 2012: 203). A smo vseeno od te norme velikokrat zaradi praktičnih razlogov odstopili.³ Rešitve v pravopisu so rojenemu govorniku velikokrat tuje in neznane. Pri samih bralcih se je v končni analizi potrdilo splošno dejstvo, da za raziskovanje govorjenega jezika ni pomembna samo kvantiteta posnetkov, ampak tudi številčnost govorcev in njihova socialno-geografska raznolikost. Bralca sta imela kljub isti pravorečni šoli na nacionalnem radiu različne glasovne značilnosti. „Tako je npr. reduciranje nenaglašanih in tudi naglašanih samoglasnikov /.../ v knjižnem govoru nezaželeno“ (Tivadar 2012: 209), a se v posnetkih pri besedah s kratkim naglašanim *a* redukcija vseeno pojavlja.⁴ Notacijo, ki smo jo od začetka uporabljali pri transkripcijah za Govorca, smo morali zaradi razkoraka med teorijo in samimi posnetki naknadno spremeniti. Vse simbole glasov, ki smo

² V slovarju izgovarjav smo najprej popravili večino glagolov. Napačno so bili naglašeni na primer predvsem deležniki moškega spola na *-l*, ki imajo predpisan naglas na korenskem samoglasniku, velikokrat so manjkale tudi dvojnice. Večkrat smo spreminjali tudi izgovore tujk in izgovor črke *l* kot *μ*.

³ Take spremembe so: uvedba ozkega izgovora *o* v prihodnjiku *bo*, čeprav je nenaglašen in bi moral biti izgovorjen široki *ɔ*, uvedena dvojnica pri glavnih števnikih na *-ajst* in na primer popravek izgovora *džójt* v *džójnt*.

⁴ O razmerju med govorjenim in pisnim jezikom več v člankih Hotimirja Tivadarja *Vzpostavitev razmerij med govorom in branjem, recitacijo in igranjem, Gradivna utemeljenost opisa slovenskega govorjenega (knjižnega) jezika, Nevarna razmerja med pisnim in govorjenim jezikom* in *Nove usmeritve pri raziskavah govora s pogledom v preteklost*. Pri delu z Govorcem je izredno pomembno, da se zavedamo nedoločenosti tega razmerja, saj se s takšnimi projekti vendarle (ne)hote vrednoti in usmerja govorjeni knjižni jezik.

jih uporabljali, z opisi in primeri položajev v besedah navajamo v Tabeli 1, saj so tudi fonetični zapisi v diplomskem delu v notaciji MRPA. Poleg pisave MRPA, ki smo jo uporabljali pri Govorcu, navajamo tudi mednarodno fonetično abecedo IPA.

Glas ⁵	MRPA	IPA ⁶	Opis glasu	Primer
A	a	a	kratki <i>a</i>	riba
A1	"a	'a	naglašeni kratki <i>a</i>	bràt
A2	"a:	'a:	dolgi naglašeni <i>a</i>	sám
B	b	b	<i>b</i>	objèm
B1	b_n	ḃ	favkalni <i>b</i>	óbme
B2	b_f	bʱ	zobnoustnični <i>b</i>	ob figi
C	ts	ts	<i>c</i>	stríc
C1	dz	dz	<i>c</i> pred nezvočnikom	Kócbek
Č	tS	tʃ	<i>č</i>	člôvek
Č1	dZ	dʒ	<i>dž</i>	džúngla
D	d	d	<i>d</i>	déd
D1	d_l	dˀ	obstranski <i>d</i>	dléto
D2	d_n	ḏ	favkalni <i>d</i>	dnò
E	"e:	'e:	dolgi ozki naglašeni <i>e</i>	Péter
E1	E	ɛ	kratki široki <i>e</i>	deréza
E2	"E	'ɛ	naglašeni široki <i>e</i>	kmèt
E3	"E:	'ɛ:	naglašeni široki dolgi <i>e</i>	têta
E4	@	ə	polglasnik	məglà
E5	"@	'ə	naglašeni polglasnik	məglà
F	f	f	<i>f</i>	fànt
G	g	g	<i>g</i>	glágol
H	x	x	<i>h</i>	mèhko
H1	G	ɣ	<i>h</i> pred zvonečim nezvočnikom	súh grm
I	i	i	kratki <i>i</i>	mámi
I1	"i	'i	kratki naglašeni <i>i</i>	sít
I2	"i:	'i:	dolgi naglašeni <i>i</i>	píla
J	j	j	<i>j</i>	jájce
K	k	k	<i>k</i>	krík
L	l	l	<i>l</i>	labílen
L1	l'	ɭ	mehčani <i>l</i>	póljski
M	m	m	<i>m</i>	imám
M1	F	ɱ	zobnoustnični <i>m</i>	simfónija
N	n	n	<i>n</i>	Nína
N1	N	ɳ	mehkonebni <i>n</i>	Ánhovo

⁵ Tabela so sestavili na Institutu „Jožef Stefan“, vsi glasovi niso v skladu z novjšimi fonetičnimi raziskavami. Glasovi so bili dani na podlagi Slovenskega pravopisa (njegov slovarski del je tudi temelj Amebisovega slovarja izgovorjav). V članku z naslovom *Normativni vidik slovenščine v 3. tisočletju – knjižna slovenščina med realnostjo in idealnostjo* avtor Hotimir Tivadar piše, da je „v slovenskem knjižnem jeziku /.../ prav trajanje eden zelo pomembnih dejavnikov govorjenega jezika, imelo naj bi tudi fonološki značaj (Toporišič 2000: 46–73), kar se sicer po nekaterih raziskavah, predvsem novjših, ne potrjuje (Srebot Rejec 1998, Šuštaršič-Komar-Petek 1995; Tivadar 2004a, b)“. Ob delu z govorno zbirko sem tudi sam potrdil, da občasno „nenaglašeni e in o niso »široki«“ (Tivadar 2010: 112). Zato smo uvedli nova znaka za ozka nenaglašena e in o. Ožino in širino teh glasov smo ocenjevali slušno, v kolikost pa v večini primerov nismo posegali, saj bi za to potrebovali obsežnejšo analizo in bi natančne meritve enostavno vzele preveč časa.

⁶ Notacija IPA je povzeta po članku *Računalniški simbolni fonetični zapis slovenskega govora* avtorjev Zemljak idr.

N2	n'	n ^j	mehčani <i>n</i>	sánjski
N3	F	ɱ	zobnoustnični <i>n</i>	informácija
O	"o:	'o:	ozki dolgi naglašeni <i>o</i>	nós
O1	O	ɔ	kratki široki <i>o</i>	vodnár
O2	"O	'ɔ	kratki naglašeni <i>o</i>	snòp
O3	"O:	'ɔ:	široki dolgi naglašeni <i>o</i>	vôda
P	p	p	<i>p</i>	pípa
P1	p_n	ɸ	favkalni <i>p</i>	rób míze
P2	p_f	p ^f	zobnoustnični <i>p</i>	Sápfo
R	r	r	<i>r</i>	tráva
S	s	s	<i>s</i>	sílos
Š	S	ʃ	š	šumíš
T	t	t	<i>t</i>	tát
T1	t_l	t ^l	obstranski <i>t</i>	tlà
T2	t_n	ɸ	favkalni <i>t</i>	tnálo
U	u	u	kratki <i>u</i>	brátu
U1	"u	'u	kratki naglašeni <i>u</i>	krùh
U2	"u:	'u:	dolgi naglašeni <i>u</i>	krúha
V	v	v	<i>v</i>	síva
V3	U	u za soglasnikom	dvoglasniški <i>u</i>	sív
L2	U	u za soglasnikom	dvoglasniški <i>u</i>	pôl
Z	z	z	<i>z</i>	níz
Ž	Z	ʒ	ž	žába
PRESLEDEK	—		presledek	
ZLOG			ločilo zlogov	ža ba

Tabela 1: Seznam glasov

3 Zgodovina sintetizatorjev govora

Želja ustvariti stroj, ki bi bil zmožen govora in prostega glasovnega odzivanja, vznemirja znanstvenike že od leta 1930, ko je Homer Dudley razvil sintetizator z imenom VODER (angl. Voice Operating Demonstration); upravljal se je z roko, nogo in prsti. Sicer je že v drugi polovici 18. stoletja Christianu Kratzensteinu uspelo z mehanično napravo s pomočjo cevi, oblikovanih po govornem traktu, izgovarjati samoglasnike. Konec 19. stoletja so razvili naprave, namenjene snemanju citatov, z njimi so v poslovnem svetu zmanjšali stroške najema stenografov. Snemalniki so delovali na principu mikrofona, zvočni pritisk je na valj, oblit z voskom, z iglo, podobno gramofonski, vtiskal zvočni signal, ki se je lahko naknadno večkrat predvajal. Leta 1952 so Davis, Biddulph in Balashek razvili sistem, ki je s formantnimi in kvantitetnimi analizami že imel sposobnost prepoznave govora posameznega govorca. V šestdesetih letih so tako sistemi poznali besedišče, veliko od 10 do 100 besed, ki

so ga prepoznavali z enostavnimi akustičnofonetičnimi lastnostmi posnetega govora. V sedemdesetih letih se je besedišče sintetizatorjev povečalo in je tako obsegalo od 100 do 1000 besed. Uvedle so se metode LPC,⁷ ena izmed njih se uporablja pri Govorcih 1 in 2. V osemdesetih letih je bilo poznavanje besed že neomejeno, prepoznavanje govora pa je temeljilo na statističnih metodah predvidevanja. Tu že govorimo o metodi PMM (prikriti modeli Markova), ki jo bomo podrobneje razložili v poglavju *Sinteza* na strani 9. V devetdesetih letih se je prepoznavanje govora dvignila na raven razumevanja in prepoznavanja skladnje v prostem govoru. Zdaj smo že na stopnji, ko je sinteza govora postala uporabna v vsakodnevni in poslovni rabi.⁸

V Sloveniji se je zanimanje za sintezo slovenščine pojavilo veliko pozneje. Vzrok je verjetno v politični situaciji v preteklosti in samem (ne)dojemanju slovenščine kot govorjenega knjižnega jezika v Jugoslaviji. Prvi z izkušnjami s sintetizatorji govora slovenskega jezika je Hribar. Razvil je postopek za samodejno pretvarjanje slovenskega besedila v fonetični prepis besedila. Razvoj sintetizatorja govora se je nadaljeval na Institutu „Jožef Stefan“ v Ljubljani.⁹ Weilguny je, podobno kot pred njim Hribar, besedilo pretvoril v fonetični prepis, mu na zelo preprost način priredil prozodične parametre in ga posredoval dvema formantnima sintetizatorjema, razvitima za angleški jezik (Weilguny 93). Raziskave na omenjenem inštitutu sta nadaljevala Dobnikar in Šef (Šef 96). Razvili so difonski sintetizator govora (Šef 98), ki so ga pred nedavnim poklonili v uporabo slepim in slabovidnim osebam (Žganec Gros 2000). Ravno ta Šefov difonski sintetizator (Govorec 1) je tisti, ki ga zdaj uporablja tudi trenutni Govorec 2. „Program s/m/o razvili na Odseku za inteligentne sisteme (E9) na Institutu „Jožef Stefan“ v sodelovanju s podjetjem Amebis in MMC RTV Slovenija“ (<http://dis.ijs.si/rtv-govorec/>, dostop 29. 4. 2013). Razvoj in raziskave s področja govornih tehnologij za slovenski jezik so še vedno dejavne. Izvajajo jih na Fakulteti za elektrotehniko Univerze v Ljubljani (Laboratorij za umetno zaznavanje, sisteme in kibernetiko), na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru, na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, na Institutu „Jožef Stefan“, na Naravoslovnotehniški fakulteti Univerze v Ljubljani in v podjetjih

⁷ Metoda LPC se uporablja pri sintetizatorjih, ki temeljijo na lepljenju govornih enot (npr. dvoglasov). Analiza in sinteza pri tej metodi potekata na podlagi linearne predikcije (angl. linear predictive coding).

⁸ O zgodovini sinteze govora smo povzemali iz Žganec Gros (2000) in B. H. Juang idr. (2005). Več o zgodovini, vrstah in lastnostih obstoječih govornih zbirk v Zemljarič Miklavčič (2008: 29–54).

⁹ Pri navajanju celega imena Instituta „Jožef Stefan“ v Ljubljani bomo uporabljali obliko *Institut*, kot je tudi uradni naziv tega inštituta, drugače pa bomo uporabljali *inštitut*, saj je to tudi primarna izbira v SP. V Gigafidi smo preverili tudi pogostost rabe oblike *institut*, ki se uporablja skoraj petkrat manj kot *inštitut*.

Masterpoint, d. o. o., Hermes Softlab, d. d., in Alpineon razvoj in raziskave, d. o. o. (Žganec Gros idr. 2003: 47). Na seznam lahko dodamo še podjetje Amebis, d. o. o.¹⁰

4 Sintetizator govora

Sinteza je umetno poustvarjanje govora, današnji cilj je, da bi bil čim podobnejši naravnemu govoru. Tu se pojavijo največji izzivi tako za jezikoslovce kot za informatike. Govoreči stroji so različni. Lahko so samodejni odzivniki, ki imajo vse besede posnete že vnaprej, predvsem so uporabni v položajih, ko gre za skladenjsko manj zahtevna besedila (v dvigalih, na železniških postajah ...). Pri obsežnejših besedilih se uporablja samodejna sinteza posameznega zapisanega besedila, kar obravnavamo tudi v diplomskem delu. Za dober sintetizator so pomembne predvsem tri lastnosti:

- kakovost sinteze,
- obseg možnih izgovorjenih besed, stavkov, intonacij,
- nezapletenost izdelave in nezahtevnost računskih ter drugih operacij.¹¹

Postopki ustvarjanja oziroma delovanja sintetizatorja govora so različni. Pri prvih je potekala sinteza govora v naslednjih fazah:

1. faza: grafemsko-fonemska pretvorba besedila,
2. faza: nastavljanje prozodičnih parametrov,
3. faza: izgradnja govorne zbirke,
4. faza: preizkus govorca.

¹⁰ Amebis, d. o. o., je podjetje, ki se ukvarja z jezikovnimi tehnologijami, cilj razvoja je predvsem čim boljše razumevanje besedila in s tem boljši samodejni lektorji, sintetizatorji govora, virtualni asistenti. Več podatkov o podjetju in njihovih projektih je dosegljivih na njihovi spletni strani (<http://www.amebis.si/>).

¹¹ V pomoč pri spoznavanju s sintetizatorji in sintezo so: Rozmanovi deli (diplomsko in magistrsko) *Lepljenje trajajočih glasov v sintezi govora* (2000) in *Sinteza govornega signala na osnovi metode HNM* (2005), monografija avtorice Žganec Gros z naslovom *Samodejno tvorjenje govora iz besedil* (2000), *Govorni korpusi* (2008) avtorice Zemljarič Miklavčič in članek Žganec Gros idr. (2003) z naslovom *Govorne tehnologije: pridobivanje in pregled govornih zbirk za slovenski jezik*.

Pri novejših postopkih sinteze je 2. faza (nastavljanje prozodičnih lastnosti) samodejna s pomočjo programske analize govornih signalov, ki transkripcijo in govorni signal vzporedno statistično primerja. Rezultati analize se pozneje ročno popravijo in izpopolnijo. Izgradnja zbirke je sicer vedno potrebna, a so njeni postopki in elementi za različne vrste potvorbe prilagojeni.

S podjetjem Amebis, d. o. o., sodelujem slabi dve leti, na začetku sem se ukvarjal predvsem z naglaševanjem in popravljanjem glagolov v slovarju izgovarjav, nato pa s preverjanjem transkripcije posnetkov iz njihove govorne baze. To delo je tudi vir in navdih diplomskega dela. Temo sem izbral, ker sem želel raziskati aplikativnost ob praktičnih izkušnjah s področja fonetike, notacije, pravorečja (in s tem tudi srečevanje s sociolingvistiko). Pri delu s sintetizatorjem govora sem počasi spoznaval tudi obsežnost področja raziskovanja svojega dela in vseh korakov, ki so za dober sintetizator potrebni:

1. korak: slovnična analiza vhodnega besedila,
2. korak: nadomeščanje simbolov, števk in kratic z besedami,
3. korak: grafemsko-fonemska pretvorba (s slovničnimi pravili in bazo znanih besed),
4. korak: določanje prozodičnih parametrov in
5. korak: sintetiziranje govornega signala (Rozman 2005: 3–4).

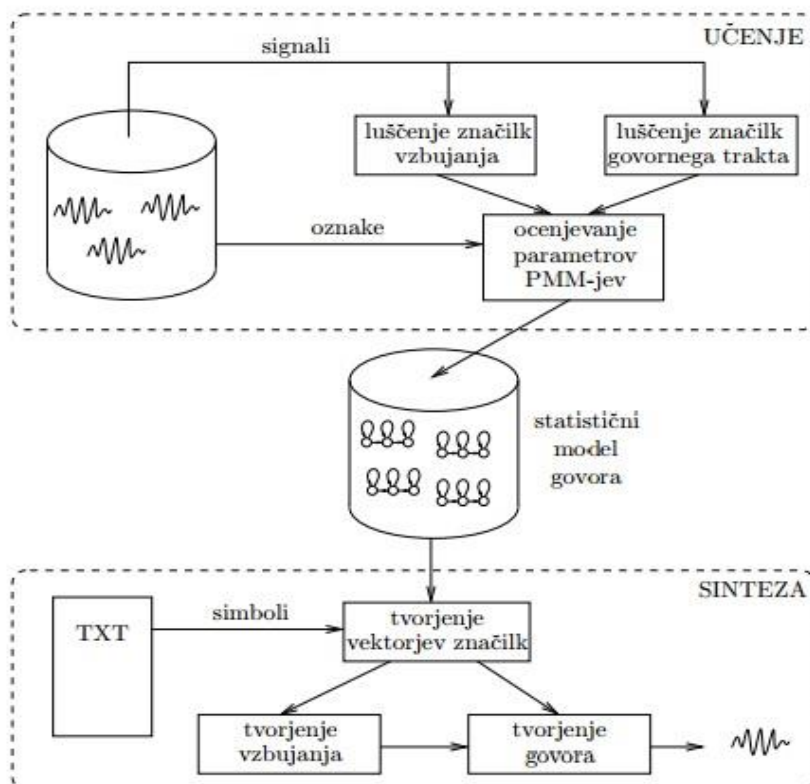
Zaradi obsežnosti teme in predvsem slovenističnega vidika na temo raziskovanja v diplomskem delu bomo delovanje sintetizatorja govora razdelili na dva dela: na obdelavo vhodnega besedila (slovnična analiza, grafemsko-fonemska pretvorba, težave ročnega popravljanja transkripcije, določanje prozodije ...) in na samo sintezo govornega signala ter lastnosti govornih zbirk, ki jih ta potrebuje.

Najprej se bomo posvetili sintezi – vrstam sinteze (čeprav je to zadnja stopnja potvorbe), saj jih lahko predstavimo na kratko, ker je to za jezikoslovca korak, na katerega nima večjega vpliva. Nato bomo v poglavju *Vhodno besedilo* (stran 8) obravnavali še prve štiri točke pretvorbe, pri katerih bomo sproti opisali tudi vse prepreke, ki so se pojavljale pri delu z Govorcem 3.

5 Sinteza

Z besedo *sinteza*, kot smo že povedali v uvodu, označujemo v tem diplomskem delu sam proces umetnega poustvarjanja človeškega govora, to je zadnja stopnja delovanja sintetizatorja govora (peti del po Rozmanovi delitvi korakov pretvorbe v umetni govor). Ker je teh več vrst, so tudi tipizacije sintez različne. „Po eni izmed delitev se postopki za sintezo govornega signala delijo v naslednje tri glavne skupine [d'Alessandro 95]: artikulatorni sintetizatorji govora, formantni sintetizatorji govora in sintetizatorji govora, ki temeljijo na združevanju osnovnih enot“ (Žganec Gros 2000: 98). Prvi so fizični modeli, zasnovani na podlagi človeškega govornega trakta, pretoka zraka in nastavljanja parametrov za tlak, napetost glasilk, položaj jezika ... Formantni sintetizator temelji na domnevi, da se lahko vpliv in obnašanje govornega trakta opišeta s formantnimi frekvencami in amplitudami, pri sintezi pa pride do umetne rekonstrukcije formantnih značilnosti. Pri drugih delitvah se znotraj tretjih sintetizatorjev po d'Alessandru pojavijo različne podvrste, razlikujejo se predvsem po osnovnih enotah, ki jih uporabljajo za združevanje (difoni,¹² besede, stavki). K novejšim metodam spada metoda PMM (prikriti modeli Markova) oziroma v angleščini HMM (angl. hidden Markov models). Tu je že govora o korpusnih sintetizatorjih. Njihova posebnost je, da imajo v primerjavi s starejšimi več posnetkov iste govorne enote, zato lahko izbirajo najprimernejšo in s tem zmanjšajo agresivno prilagajanje govornih signalov, kot se to dogaja pri Govorcu 1, ki ima v bazi samo po en primer enega difona, zato mora za sestavitev vseh v eno besedo veliko bolj zgladiti signale elementov (osnovni ton in moč), kot bi jih bilo treba, če bi imeli pri postopku sinteze možnost izbirati različne posnetke istih difonov. Primernost bi se določala po podobnosti končnega signala enega difona z začetkom drugega. „Postopek sinteze govora z uporabo PMM-jev se razlikuje od bolj razširjenih postopkov v tem, da ogrodje PMM-jev ne uporablja zgolj za segmentacijo in označevanje govorne zbirke, pač pa gre še korak naprej in ga uporablja tudi kot model za tvorjenje govora“ (Vesnicer idr. 2004). Slika 1 iz tega članka prikazuje proces sinteze po metodi PMM.

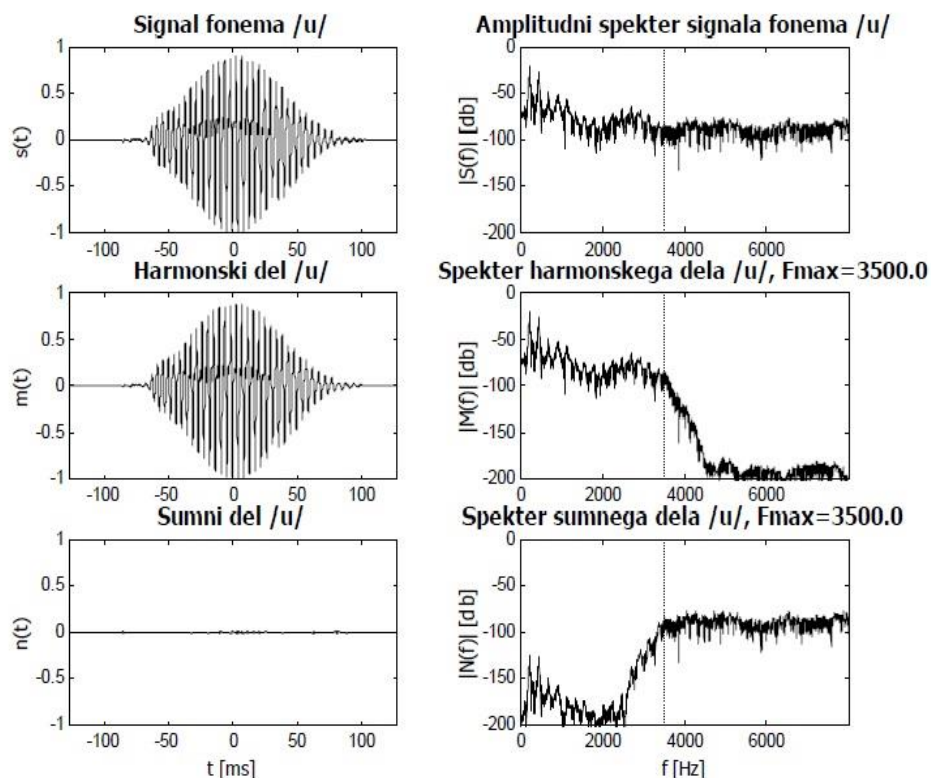
¹² Difon oziroma dvoglasnik je kombinacija dveh glasov, ki sta izgovorjena eden za drugim. Kot primer so na strani 14 navedeni difoni besede *lipa*, ti so: *_l*, *l''i*:, *''ip*, *pa*, *a_* (podčrtaj nakazuje tišino).



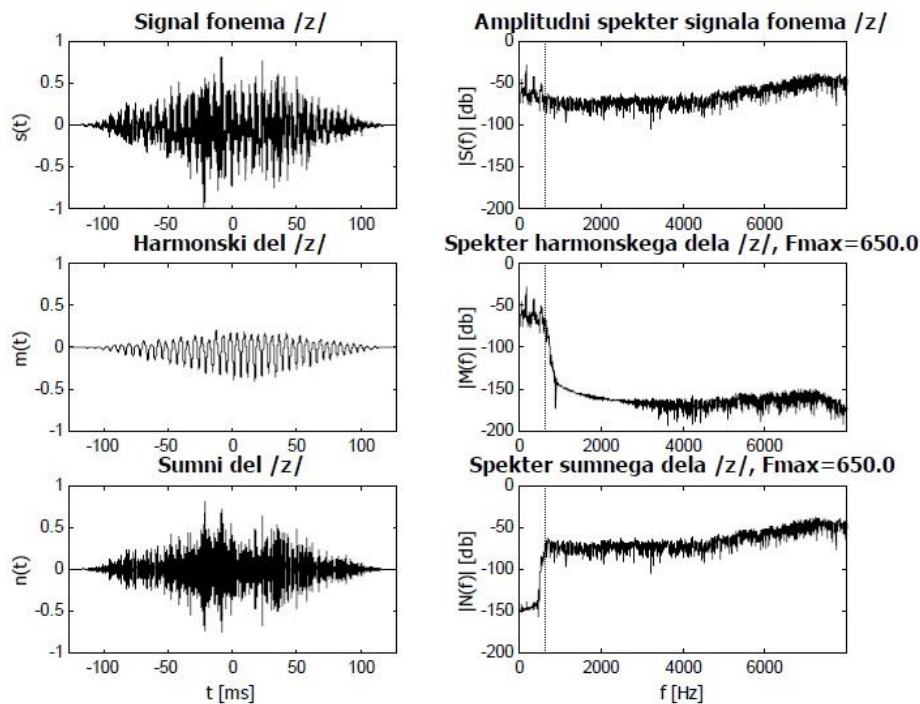
Slika 1: Shematski prikaz postopka učenja in sinteze govora z uporabo prikritih modelov Markova (Vesnicher idr. 2004)

Ena izmed zadnjih metod sinteze je metoda HNM (angl. harmonic plus noise modulation), je izpeljanka sinusne metode. „Govorni signal dobimo tako, da vzbujevalno funkcijo, ki ponazarja periodične impulze glasilk, filtriramo s filtrom, ki modelira obliko govorne cevi“ (Rozman 2005: 5). O harmonski sintezi govorimo, „/k/adar za vzbujevalno funkcijo namesto poljubno razporejenih frekvenc uporabimo osnovno frekvenco in njene harmonske frekvence“ (Rozman 2005: 7). Ker pa ta zelo dobro deluje samo na zvonečih glasovih, se lahko uporablja hibridno skupaj z generiranjem šuma, saj se pri „/p/ri zvonečih in tudi nezvonečih glasovih /.../ v posameznih delih pojavlja šum kot posledica zračnih turbulenc v govorni cevi“ (Rozman 2005: 8). Metoda HNM je tako skupek harmonskih sinusnih sintez in sinusnih sintez s šumom. „Bistvo metode je dekompozicija govornega signala na harmonski del, ki ga modeliramo s harmonsko sinusno sintezo, in šumni del, ki ga modeliramo s filtriranjem šuma“ (Rozman 2005: 9). Sliki 2 in 3 iz Rozmanove *Sinteze govornega signala na osnovi metode HNM* tudi za jezikoslovce jasneje nakažeta bistvo zapletenega opisa. Gre za dekompoziciji glasov u in z na harmonski in šumni del. Dekompozicija je analizni del procesa sinteze, ki izračuna harmonski in šumni del in vse

koeficiente iz posnetka človeškega govora, torej vse podatke, ki so potrebni za poznejšo uresničitev umetnega govora. Ker pri izgovoru glasu u ne prihaja do turbulenc v govorni cevi, je šumni del glasu zanemarljiv, pri glasu z pa zaradi zvenečnosti pričakovano vidimo tudi harmonski del, a večino energije pri signalu predstavlja šumni del.



Slika 2: Dekompozicija glasu u na harmonski in šumni del (Rozman 2005: 10)



Slika 3: Dekompozicija glasu z na harmonski in šumni del (Rozman 2005: 11)

Po seznanitvi z večino sintez lahko zaključimo, da je razdelitev sintetizatorjev po d'Alessandru preozka; tako lahko sintetizatorje razdelimo na:

- a) artikulatorne,
- b) formantne,
- c) sintetizatorje, ki združujejo enote (npr. difonski),
- č) sinusne (varianta sinusnih je tudi metoda HNM),
- d) tiste, ki temeljijo na metodi HMM oz. modelih Markova.

Druga možna delitev, ki bi bila slovenistom tudi veliko bližje, saj je glavni del našega dela pri vходу v sintetizator govora, bi se osredotočala na različne vrste vhodnega besedila. Če gre za omejene možnosti branja, ponavadi le za lepljenje vnaprej posnetih stavkov, bi govorili o zelo enostavnih in precej omejenih, specializiranih sintetizatorjih. Kadar pa gre za t. i. „text to speech“, ali poslovenjeno „prenos besedila v govor“ sintetizatorje, poimenujejo jih tudi korpusni sintetizatorji, bi govorili o splošnih ali samodejnih sintetizatorjih, ki

besedišča ne morejo predvideti in zato za delovanje potrebujejo velike slovarje in govorne zbirke.

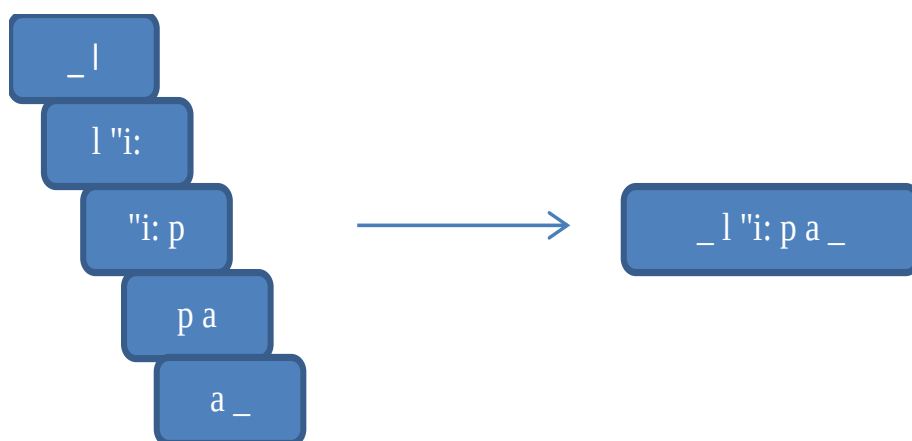
5.2 Sintetizator govora Govorec

Govorec je ime sintetizatorja govora, ki so ga sprva razvili na Institutu „Jožef Stefan“ v Ljubljani. Trenutno obstajajo tri verzije Govorca. Drugo je razvil Amebis, d. o. o., sam pa sem se pri delu v podjetju ukvarjal z Govorcem 3, ki bo deloval z novo vrsto sinteze (metoda HNM namesto TD-PSOLE) in novo govorno zbirko (več ur posnetkov govorjenega besedila namesto baze difonov); prav tako ga razvija Amebis, d. o. o. Govorca 1 in 2 pa se razlikujeta v tem, da Govorec 2 uporablja novo bazo besed oziroma slovar izgovarjav, ki je veliko obširnejši in bolj izpopolnjen. Ker se pri pojasnjevanju temeljev delovanja Govorcev 1, 2 in 3 ne moremo izogniti terminu govorna zbirka, bomo v tem poglavju opisali še tega.

5.2.1 Sintetizatorja Govorca 1 in 2

Govorca 1 in 2 delujeta po postopku TD-PSOLA, sta difonska sintetizatorja – torej lepita difone. Sintetizatorji lahko lepijo tudi posamezne foneme, trifone, besede, stavke ali pa kar cele povedi. „Ta postopek enostavno zapišemo kot matematično vsoto zamaknjenih signalov fonemov. Signale zamaknemo tako, da se rahlo prekrivajo /.../ Signale fonemov, ki jih lepimo skupaj, moramo vnaprej posneti “ (Rozman 2000: 8). Ker govorci ne morejo biti toliko konstantni v izgovoru, so difoni izgovorjeni z različnimi frekvencami (osnovni ton) in jakostmi (energija), zato je treba posnetke dodatno obdelati. Govor bo tako sicer nenaraven, slišati bo kot sestavljanje različnih glasov. Frekvence vseh enot moramo izenačiti, potem tona zvežemo na stiku in nato prevzorčimo še celo besedo, da dosežemo počasno spreminjanje izgovora celotne besede, kot se to zgodi pri naravnem govoru. Tu naletimo na pojem koartikolacija. „Ker naši govorni organi niso sposobni nenadnih in velikih sprememb v gibanju, se to odraža na akustični realizaciji posameznega fona, ki je zelo odvisna od pozicije govoril pred in po izgovarjavi le-tega“ (Rozman 2004). Po izravnavi frekvence sledi še izravnava energije. Beseda ima največ energije na naglašenem glasu (približno dvakrat več). Seveda je ta postopek še veliko bolj zapleten, kot je razloženo tukaj, potrebni so še filtriranje, računanje period, uporaba logaritmov, manipuliranje faze in posebne enačbe. Govorca 1 in 2

pa frekvenčno in fazno ne obdelujeta trajajočih nezvonečih glasov, saj ti nimajo periode in s tem osnovne frekvence.¹³ Ti se po mejah usklajujejo le po energiji. Večje težave se pojavijo pri glasu *r*, saj ima ta zelo nestalno periodo. Sintetizatorji govora večinoma lepijo posnetke iz govornih baz. Če je npr. vsak difon besede *lipa* (*_l*, *l"i:*, *"i: p*, *p a*, *a _*) element v govorni bazi (glasove morajo predhodno posneti posamezniki), je shema sinteze besede *lipa* takšna:

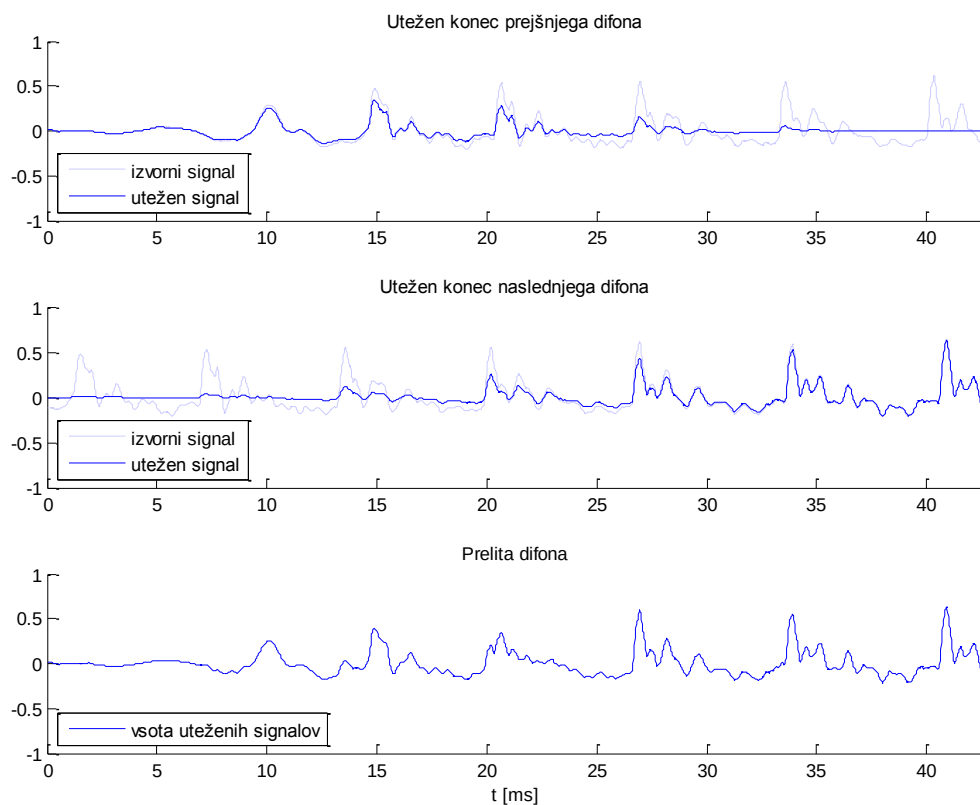


Slika 4: Shema sinteze besede *lipa* po postopku TD-PSOLA

Pri lepljenju elementov (npr. difonov) se morata usklajati ton in energija, saj je nemogoče vse glasove posneti z isto močjo in višino tona. Ton (f_0 oz. osnovna frekvenca) dveh elementov (npr. *l* in *l"i:*) bi se izenačil zvezno in počasi, kot se tudi pri človeškem glasu. Pomembni besedi pri lepljenju elementov sta *prehod* in *prelivanje*.¹⁴ V difonih se namreč prehodi med glasovoma (npr. glasovoma *l* in *"i:*) ne spreminjajo, energija in frekvenca se ne prilagajata. Zlivanje pa se zgodi med končnim delom prvega difona in začetnim delom drugega difona, v primeru lepljenja difonov *l* in *l"i:* se v procesu prelijeta glasova *l* iz obeh difonov, notranji prehodi v difonih pa se ne spreminjajo. Postopek lepljenja je računanje matematične utežene vsote govornih signalov, ki je viden na Sliki 5.

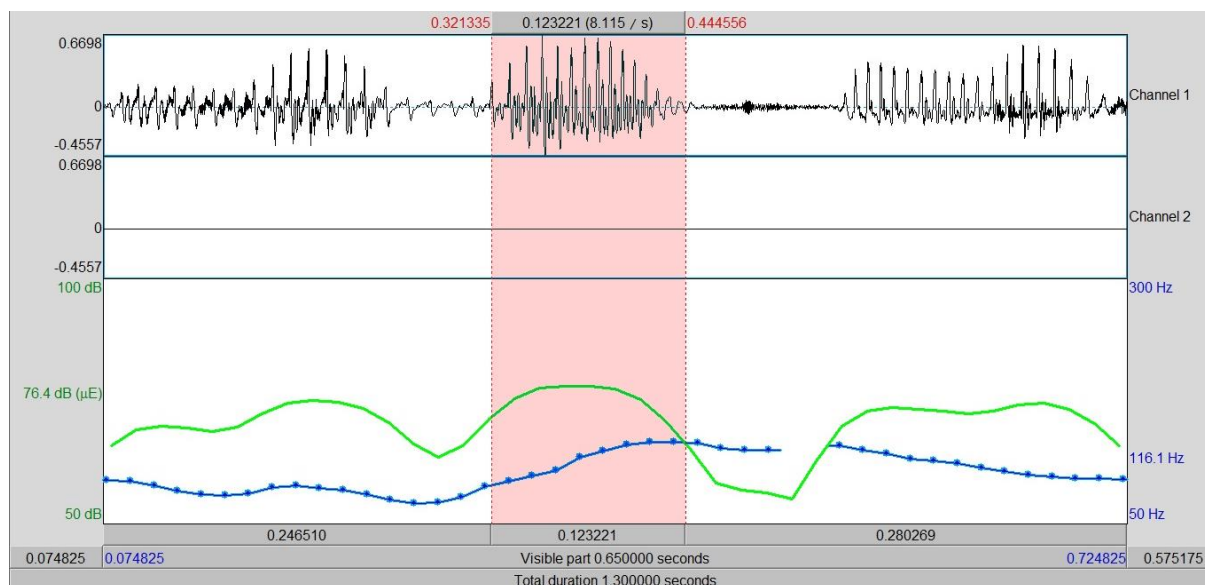
¹³ Osnovno frekvenco glasu smo iskali s programom Praat in F_0 .

¹⁴ Pri besedi *prelivanje* gre za slovenski prevod angleške besede „crossfade“, pogosta v krogih avdioinženiringa.



Slika 5: Prelivanje difonov

Naglas besede in osnovni ton besede prikazuje analiza besede *zavóhala* (Slika 6).



Slika 6: Naglas in analiza besede *zavóhala*

Označeni del signala je ozki in naglašeni ó. Iz okna programa Praat se dobro vidi, da sta na mestu naglasa zvišana ton (modra črta) in jakost (zelena črta). Prav to mora potem sintetizator tudi posnemati pri sintezi. Glasove zveže na mikroravni in na makroravni in ponastavi ton ter jakost, da ima beseda naglas na pravem mestu. Difonski sintetizatorji potrebujejo zelo obsežno bazo posnetkov, saj morajo imeti za sintezo besedila vse možne difone v jeziku, ki jih nato lepijo skupaj. Govorec 1 (in s tem tudi 2) pozna 1152 dvoglasov. Vsak difon je posnet le enkrat, tako ni mogoče iskati najprimernejšega, prilagajanje na stikih je agresivnejše, posledica pa slabša sinteza. Število potrebnih difonov za možnost sinteze je odvisno od samega jezika, španščina jih potrebuje npr. 800 in nemščina okoli 2500. Zato je dobro, da je govorna baza velika in izbira pestra.

5.2.3 Sintetizator Govorec 3

Za Govorca 3 so pri Amebisu, d. o. o., ustvarili govorno zbirko, ki sta jo posnela profesionalna napovedovalca z nacionalnega radia, oba naj bi predstavljala normo oziroma standard današnjega pravorečja. Gre za dvajset ur posnetkov, ki so razdeljeni na posamezne odseke, ponavadi so to posamezne povedi. Te povedi so vzete iz korpusa besedil in so izbrane tako, da je končni rezultat čim širši izbor kombinacij glasov, možnih intonacij, podredij, besednih vrst (npr. medmeti) ... Teh dvajset ur posnetkov (deset ur na enega bralca) je razdeljenih na okoli 4000 delov pri vsakem bralcu (del je večinoma poved). Poleg posnetka govora je posnet še signal utripanja glasilk, kar olajša analizo (predvsem zvonečnosti in nezvonečnosti); govorimo o laringografskem posnetku utripanja glasilk. Ker so se posnetki delali večkrat, so med njimi slišne razlike. Napovedovalca se proti koncu posamezne seje bolj motita, bereta hitreje ali počasneje, spreminjata besedilo in požirata glasove. Na posnetkih se pojavlja tudi šum listov, kar lahko poznejšo računalniško analizo moti, saj je šum papirja zelo podoben šumu človeškega govora. Ta šum pa je pri analizi HNM, ki jo Govorec 3 uporablja, izredno pomemben, zato je tu nevarnost napačne samodejne analize (šuštenje listov je lahko dojeto kot glas). Prednost Govorca 3 bo predvsem široka možna izbira različnih glasov, besed ali pa celo kar celih stavkov. Govorec 3 bo v govorni zbirki, ki so jo posneli, iskal največje možne dele posnetkov, ki jih bo izbral iz baze. Če bo teh možnosti več, bo izbral najprimernejše, takšne, ki bodo potrebovali na prehodih najmanj korekcije. Ker se je pri transkripcijah označevala tudi makroprozodija, bo tako naravnost govora še večja. Pričakuje

se (saj še ni bil testiran), da se bo npr. izboljšala tudi intonacija (vprašalna, povedna, vzklična). Naravnejši bi morali biti tudi prehodi pri različnih stavčnih podredjih ali priredjih. Naravnejši bodo odmori. Na posnetkih so v transkripciji označeni in ohranjeni tudi kašlji, brundanje, razna mašila, ki se bodo lahko nato naključno pojavljala med sintezo, da bo ta še bolj človeška.

5.2.4 Govorne zbirke

Govorne zbirke so obvezen del tako pri današnjih jezikoslovnih raziskavah kot pri ustvarjanju umetnega govora. Zbirke posnetkov so zelo različne. Predvsem pa se njihove lastnosti oblikujejo glede na namen. Pri Govorcu 3 je bilo nujno dobiti čim več možnih kombinacij besed in intonacij dveh bralcev, saj bo lahko uporabnik izbral med dvema bralcema, in sicer moškim in ženskim glasom. Pomembno je bilo tudi, da obvladata zborni in knjižni pogovorni jezik, ki naj bi bil trenutno standarden. Čeprav se pri tem med slovenisti občasno lomijo različna mnenja, kako bi moralo biti in ni, sta izbrana bralca trenutno vendarle reprezentativna predstavnika norme.¹⁵ V zbirki so posnete tudi besede, kot so pika, vejica in evro, saj je Govorec uporaben predvsem za slabovidne, ki potrebujejo popolno informacijo o branem besedilu. Govorno zbirko opredeljujejo posnetki govornih signalov, priložene označitve in transkripcije. Pod oznake spadajo tudi podatki o govornikih (spol, starost, poklic, izobrazba ...). „Najpogostejši so mikrofonski posnetki zvočnih signalov, ki se pojavljajo med govorjenjem, in laringografski posnetki vibriranja glasilk v grlu. Uporabljajo se tudi vizualni in rentgenski posnetki govoril. Prvi se nanašajo zgolj na slikovne posnetke obraza, predvsem ustnic, drugi pa na globinske slike prereza celotnega sistema govoril. V zadnjem času se pogosto analizirajo tudi posnetki biomedicinskih signalov, ki se pojavljajo v govorilih in tudi človeških možganih“ (Žganec Gros idr. 2003: 49).

„Gradnja /govornega/ korpusa mora biti skrbno načrtovana, določajo pa jo poleg namena korpusa tudi finančne zmožnosti in razpoložljivi človeški viri (raziskovalci, zbiralci gradiva, transkriptorji, označevalci itd.) /.../ Pomembno je, da se načrtovalci korpusa zavedajo, da je gradnja korpusa ciklični proces, ki ob sprotnem evalviranju lahko zahteva spreminjanje začetnih izhodišč, pa tudi proces, ki se v idealnih okoliščinah ne bi nikoli

¹⁵ Več o standardnem in knjižnem jeziku v drugem poglavju na strani 2.

končal, saj bi morali korpus nekončno dopolnjevati in konsolidirati“ (Zemljarič Miklavčič 2008: 93). Govorne zbirke, ki zadevajo fonetično-fonološke raziskave, in tiste za potrebe govornih tehnologij so povsem drugačne od t. i. referenčnih korpusov, ki vsebujejo tudi transkripcije. Ti so namenjeni raziskavi spontanega govornega jezika. Raziskujejo se predvsem slovnično-leksikalne lastnosti jezika. Zato so govorne zbirke tudi čisto drugače zgrajene. V vseh govornih zbirkah sta najpomembnejši reprezentativnost in uravnoteženost zbranih posnetkov. Pri fonetično-fonoloških raziskavah so pomembne pester kombinacije fonov, fonemov, difonov, besed in intonacij, pri referenčnih korpusih pa čim bolj raznoliki viri (uradni pogovori, zasebni pogovori, nastopi ...) in govorci (starost, socialni položaj, spol ...). Če gre za besedila, je zelo pomembna različnost besedilnih vrst. Z naraščanjem velikosti korpusa narašča tudi splošnost.

5.2.5 Slovenske govorne zbirke

V Sloveniji je o govornih zbirkah izdala monografijo Jana Zemljarič Miklavčič, iz katere povzemamo skrajšan seznam in opis nekaterih slovenskih govornih zbirk, ki že obstajajo. Seznam je dopolnjen še z novejšim korpusom GOS (korpus **GO**vorjene Slove**Š**čine).

- Baza izgovorjav SNABI
 - Izdelana na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru;
 - 128 govorcev;
 - vsak govorec izgovori povprečno 200 stavkov, 80 izoliranih besed, številske nize in abecedo;
 - členjena in fonetično transkribirana.
- Podatkovna zbirka GOPOLIS
 - Izdelana na Fakulteti za elektrotehniko Univerze v Ljubljani;
 - narejena za razvoj sistema za razpoznavo govora in krmilnika dialoga pri govornih poizvedbah o letalskih informacijah;
 - posnetih 25 moških in 25 žensk;

- nima urejenih avtorskih pravic.
- SPEECH-DAT
 - Evropski projekt;
 - za razvoj sistemov telefonskega govornega dialoga za delo v realnem okolju;
 - 1000 govorcev.
- POLIDAT
 - Tudi razvit za avtomatski telefonski sistem GOVORNEGA DIALOGA;
 - izbranih 1000 govorcev;
 - brana vnaprej pripravljena besedila;
 - majhen delež spontanega govora (Zemljarič Miklavčič 2008: 52–53).
- Govorni korpus GOS
 - Označujejo ga kot korpus govorne slovenščine;
 - 120 ur posnetkov (po)govora v najrazličnejših situacijah (radijske in televizijske oddaje, zasebni pogovori ...);
 - zapis govora na posnetkih je narejen v dveh različicah, standardizirani in pogovorni, ter obsega več kot milijon besed;
 - namenjen vsem, ki želijo raziskovati govorno slovenščino (<http://www.korpus-gos.net/Support/About>, dostop 4. 5. 2013).

Govorne baze se ne zbirajo le za sintetizatorje, ampak tudi razpoznavalnike govora. Govorna baza SNABI je bila prva dostopna baza za razpoznavanje govora. „S stališča kompleksnosti razpoznavalnikov govora so najobsežnejši jezikovni viri potrebni za razpoznavalnike tekočega govora. Če želimo pravilno oceniti parametre akustičnih in jezikovnih modelov, potrebujemo za izpeljavo postopka učenja velike količine učnih podatkov“ (Žgank idr. 2006). O slovenskem jezikovnem viru SloParl, ki ga sestavljata govorni in besedilni korpus, pišejo Žgank in drugi v članku z naslovom *Slovenska govorna in tekstovna baza parlamentarnih razprav za avtomatsko razpoznavanje govora*.

6 Vhodno besedilo

Vhodno besedilo je katero koli besedilo, ki mora biti strojno prebrano. Najprej je treba to besedilo obdelati in prilagoditi. Ker se v nadaljevanju dela velikokrat sklicujemo na slovar izgovarjav, ga bomo opisali pred postopki analize samega besedila.

6.1 Slovar izgovarjav kot del Govorca

Slovar izgovarjav je pri sintetizatorju slovenščine nekakšno nujno zlo. Trenutno obsega približno 6.900.000 oblik besed z določenimi MSD-ji (to so oblikoskladenjske oznake) in naglasom. Zaradi zapletenosti jezika sta samo naglaševanje in grafemsko-fonemska pretvorba veliko težja kot npr. pri francoščini ali italijanščini. Lema je po Presisovem večjezičnem slovarju razložena kot osnovna oblika. Tako se v slovarju delijo skupine besed na leme, ena lema pa označuje vse oblike določene besede znotraj iste besedne vrste. Primer: samostalnik *ljubezen* ima svojo lemo in pridevnik *ljubezenski* svojo.

ljubeški+P*→	Pvosmm-----→	ljubeških→	lju bé ških→
ljubeški+P*→	Pvosmo-----→	ljubeškimi→	lju bé ški mi→
ljubezen+Soz*→	Sozei---e-----→	ljubezen→	lju bé zen→
ljubezen+Soz*→	Sozer---e-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozed---e-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozet---e-----→	ljubezen→	lju bé zen→
ljubezen+Soz*→	Sozem---e-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozeo---e-----→	ljubeznijo→	lju béz ni jo→
ljubezen+Soz*→	Sozdi---d-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozdi---d-----g→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozdr---d-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozdd---d-----→	ljubeznima→	lju béz ni ma→
ljubezen+Soz*→	Sozdt---d-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozdt---d-----g→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozdm---d-----→	ljubeznih→	lju béz nih→
ljubezen+Soz*→	Sozdo---d-----→	ljubeznima→	lju béz ni ma→
ljubezen+Soz*→	Sozmi---m-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozmr---m-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozmd---m-----→	ljubeznim→	lju béz nim→
ljubezen+Soz*→	Sozmt---m-----→	ljubezni→	lju béz ni→
ljubezen+Soz*→	Sozmm---m-----→	ljubeznih→	lju béz nih→
ljubezen+Soz*→	Sozmo---m-----→	ljubeznimi→	lju béz ni mi→
ljubezenski+P*→	Pvomeid-----→	ljubezenski→	lju bé zen ski→
ljubezenski+P*→	Pvomer-----→	ljubezenskega→	lju bé zen ske ga→
ljubezenski+P*→	Pvomed-----→	ljubezenskemu→	lju bé zen ske mu→
ljubezenski+P*→	Pvomet---d-----→	ljubezenskega→	lju bé zen ske ga→
ljubezenski+P*→	Pvometd-n-----→	ljubezenski→	lju bé zen ski→
ljubezenski+P*→	Pvomem-----→	ljubezenskem→	lju bé zen skem→
ljubezenski+P*→	Pvomeo-----→	ljubezenskim→	lju bé zen skim→
ljubezenski+P*→	Pvomdi-----→	ljubezenska→	lju bé zen ska→

Slika 7: Slovar izgovarjav (delovno okno)

V MSD-jih (angl. morpho-lexical descriptor), ki jih vidimo v drugem stolpcu delovnega okna, so besede označene z besednimi vrstami in podrobnimi podklasifikacijami, nekaj jih naštevamo kot primere za lažjo predstavbo delovanja slovnične analize:

- samostalnik (vrsta {naselje, občno ime, lastno ime, voda, podjetje ...}, spol, sklon, število, določnost, živost, posebej pomensko število in navadno število);
- glagol (vrsta {polnopomenski, naklonski, vezni}, glagolska oblika {povednik, velelnik, pogojnik, nedoločnik ...}, čas, oseba, število, spol, način {tvornik, trpnik}, nikalnost, vrsta pomožnika {biti, imeti} in vid);
- pridevnik (stopnje, določnost ...);
- prislov (krajevni {kam, kje}, načinovni, kratnostni, merski, pogostnostni);
- zaimek;
- predlog (določen sklon in število, ki ga predlog zahteva ...);
- veznik (priredni, podredni, notranji in vprašalnice);
- členek;
- člen;
- medmet;
- števnik;
- okrajšava;
- ločilo;
- napake (napačen vrstni red, sumljiva oblika, premena odveč, sporna *-lec* in *-vec*, zatipkana oblika, množina namesto dvojine ...).

V slovarju izgovarjav se uporablja pet stolpcev. Kot primer izpostavljamo iz slovarja besedo *solata* v imenovalniku množine.

solata+Soz*	Sozmi---m-----	solate	so lá te
-------------	----------------	--------	----------

V prvem je osnovna oblika, v drugem oznaka MSD, v tretjem je beseda z vsemi oblikoslovnimi lastnostmi, v četrtem pa je beseda še naglašena. Če besede grafemsko-fonemska pretvorba ne pretvori pravilno ali da gre za kakšno glasovno izjemo znotraj pravil, je treba to zapisati v peti stolpec, vanj se zapiše pravilni fonetični zapis v pisavi MRPA, tako Govorec uporabi za branje ta zapis namesto uporabe pravil. Težave četrtega in petega stolpca so predvsem v dolgotrajnem in ročnem delu vnašanja.

6.1.2 Slovar izgovarjav in Govorec 3

Peti stolpec (fonetični zapis) v slovarju izgovarjav se je kot pomemben in nujno potreben korektiv izkazal predvsem pri nenaglašeni besedah. Prihodnjik biti *bo* je nenaglašen. Ker je nenaglašen, ga grafemsko-fonemska pretvorba zapiše v [bO]. Tako nenaglašeni *bo*, čeprav naj bi bil ozek (tako ga bereta tudi napovedovalca), sintetizator bere s širokim o.¹⁶ Vse te oblike smo morali vnašati v slovar posebej v peti stolpec. Slovar ima tudi možnost vnašanja dvojnic. Besede se zapišejo ena za drugo, a se vseeno vedno razvrstijo hierarhično, saj slovar primarno izbere obliko na prvem mestu. Takšen je npr. glagol *plesati*, ki potrebuje dvojnice pri velelniku in deležniku.

plesati+Gp*→	Gpn-----n----→	plesati→	plé sa ti,ple sá ti→
plesati+Gp*→	Gppspe--n-----n----→	plešem→	plé šem→
plesati+Gp*→	Gppspd--n-----n----→	pleševa→	plé še va→
plesati+Gp*→	Gppspm--n-----n----→	plešemo→	plé še mo→
plesati+Gp*→	Gppsde--n-----n----→	plešeš→	plé šeš→
plesati+Gp*→	Gppsdd--n-----n----→	plešeta→	plé še ta→
plesati+Gp*→	Gppsdm--n-----n----→	plešete→	plé še te→
plesati+Gp*→	Gppste--n-----n----→	pleše→	plé še→
plesati+Gp*→	Gppstd--n-----n----→	plešeta→	plé še ta→
plesati+Gp*→	Gppstm--n-----n----→	plešejo→	plé še jo→
plesati+Gp*→	Gpvspd-----n----→	plešiva→	plé ši va,ple ší va→
plesati+Gp*→	Gpvspm-----n----→	plešimo→	plé ši mo,ple ší mo→
plesati+Gp*→	Gpvjde-----n----→	pleši→	plé ši→
plesati+Gp*→	Gpvjdd-----n----→	plešita→	plé ši ta,ple ší ta→
plesati+Gp*→	Gpvjdm-----n----→	plešite→	plé ši te,ple ší te→
plesati+Gp*→	Gpdr-emt-----n----→	plesal→	plé sa →
plesati+Gp*→	Gpdr-dmt-----n----→	plesala→	ple sá la,plé sa la→
plesati+Gp*→	Gpdr-mmt-----n----→	plesali→	ple sá li,plé sa li→
plesati+Gp*→	Gpdr-ezt-----n----→	plesala→	ple sá la,plé sa la→
plesati+Gp*→	Gpdr-dzt-----n----→	plesali→	ple sá li,plé sa li→
plesati+Gp*→	Gpdr-mzt-----n----→	plesale→	ple sá le,plé sa le→
plesati+Gp*→	Gpdr-est-----n----→	plesalo→	ple sá lo,plé sa lo→
plesati+Gp*→	Gpdr-dst-----n----→	plesali→	ple sá li,plé sa li→
plesati+Gp*→	Gpdr-mst-----n----→	plesala→	ple sá la,plé sa la→
plesati+Gp*→	Gpm-----n----→	plesat→	plé sat→

Slika 8: Slovar izgovarjav (primer dvojnic)

Seveda se pojavljajo tudi trojnice in četvernice, a so, če besede niso pogosto v uporabi, nesmiselne, saj bo Govorec še vedno izbral prvo obliko. Do zdaj se je v slovarju izgovarjav naglaševalo po pravopisu, razen če raba v širšem deležu populacije ne odstopa od

¹⁶ Neskladnosti pravil in drugih rešitev v Slovenskem pravopisu z rabo so podrobneje opisane v članku *Podoba govorjenega slovenskega knjižnega jezika v Slovenskem pravopisu 2001*, ki sta ga napisala Hotimir Tivadar in Peter Jurgec.

norme. To se je dobro opazilo, ko smo po SP zamenjali vrstni red naglašanih oblik pri številih od dvanajst do dvajset. Po posvetu s sodelavci na Amebisu, d. o. o., smo dvojnice *dvánajst* in *dvanájst* pustili, saj jih ima tudi SSKJ. Tako smo naredili kompromis med SP in SSKJ, obdržali smo obe obliki in na prvo mesto zapisali *dvanájst*. Ko je to novo pravilo začel uporabljati Govorec 2, so se uporabniki pritožili. Oblika *dvanájst* se je zdela večini napačna in tuja. Vrstni red oblik smo zamenjali in dali kot primarno izbiro zopet *dvánajst*, a te v SP sploh ni. Velikokrat sta tudi bralca naglaševala besede drugače, kot so naglašene v SP. Ta težava se je pojavljala predvsem pri tujkah, ki so v SP največkrat normirane z ozkima *é* in *ó*, kar pa se v rabi ne izkazuje kot najboljše in se izgovor tujk (sploh angleških besed) v večini primerov bliža izgovoru v originalnem jeziku (zopet angleščini). Takšen primer je angleška beseda *software*, ki ima v SP osnovno obliko poslovenjeno v *softver* in je naglašena kot [s"o:ftv"E:r]. Prvi samoglasnik je pričakovano naglašen ozko, drugi pa v nasprotju z navado široko. Ker sta naša bralca vedno brala prvi *o* široko, saj je to tudi širša raba, smo izgovor zapisali kot [s"O:ftv"E:r], in tako odstopili od norme Slovenskega pravopisa. V Slovarju novejšega besedja slovenskega jezika je izgovor te besede enak našemu (*sôftvêr*).

V slovarju izgovarjav je še vedno veliko neoznačenih besed, zato je za njegovo dopolnitev potrebno še veliko dela. Pomembna bo sistematika naglasnih tipov, ki so zapisani v prvi knjigi Slovarja slovenskega knjižnega jezika, a bo ta pravila treba preveriti še na testnih besedilih in pa seveda uskladiti z informatiki, ali jih je sploh možno zapisati v obliki pravil, ki bi jih potem uporabili v slovarju izgovarjav. Slovar, ki je za zdaj še pomanjkljiv, ima za slovenščino veliko vrednost. Jezikoslovci se bodo morali v popravke, če slovar ne bo več namenjen samo za rabo Govorca, resneje vključiti. S tem seveda apeliramo tudi na poenotenje pravorečja in posledično normiranje, saj smo se morali pri delu velikokrat usklajevati med slovarji, rabo in svojim občutkom za jezik.

6. 2 Slovnična analiza pri Govorcih 2 in 3

Sintetizator mora v besedilu najprej opraviti slovnično analizo. Pri Govorcih 2 in 3 se vsem besedam določijo leme, besedna vrsta, sklon, spol, število idr. iz slovarja izgovarjav. Določijo se tudi stavčna vloga besed in skladenjske lastnosti vsake povedi. Analiza temelji na

podlagi glagolskih predlog. „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 pomene, pri čemer so pripisane tudi potrebne stilne oznake /.../ Predloge določijo tudi 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 samostalnik »kosilo« povezana tako, da je kosilo tipično predmet v tožilniku“ (Holozan 2004). Tako je Peter Holozan pojasnjeval del analize pri strojnem prevajalniku Presis, ki uporablja isto analizo kot Govorca 2 in 3.

6.2.1 Slovnična analiza in transkripcije pri Govorcu 3

Slovnična analiza je zelo obširna tema, zato navajamo le nekaj primerov, ki smo jih obravnavali. Pri popravljanju transkripcije smo lahko opazovali, kakšne napake se pojavljajo pri analizi in katere besede so napačno naglašene v slovarju, saj so vsa besedila prej obdelali, kot se obdelajo za samo sintezo. Iz napake smo morali najprej ugotoviti, ali je napaka v naglasu ali je analizator npr. izbral napačno besedno vrsto. V razpredelnici je navedenih nekaj napačnih analiz Govorca in napaka, ki smo jo predvideli in bi bila možna. Ker se te napake še preverjajo, popravki niso nujno pravilni.

Poved	MRPA	Verjetna napaka
Najemi sobo, govori s točajem in profesorjem Knuthom.	na j"E: mi s"O: bO gO vO r"i : s tO tS"a: j@m in' prO f"e: sOr j@m kn"u: tOm	Namesto velelnne oblike glagola <i>govôri</i> uporabi povedno obliko <i>govorí</i> .
Albert je bil netakten, žaljiv, porogljiv in celo grob.	"a:l b@rt jE b"i:U n"E t"a:k t@n _ Za lj"i:U _ pO rO glj"i:U _ in' tsE l"o: gr"Op	Namesto pridevnika <i>grób</i> je besedo prepoznal kot samostalnik <i>gròb</i> .
Nastopil bo lahko vsak, ki ima drsalke in ki ve, kaj je pingpong.	na st"o: piU bo lax k"o: Us"a:k(STA:rp) ki i m"a: d@r s"a:l kE _ in' ki v"e: k"a:j i"e : p"i:Nk p"O:Nk	Glagol biti <i>je</i> zamenja z glagolom jesti <i>jé</i> .
Bolezni rastlin je seveda več, od napada plesni do ličink, uši, mušic in pršic do rje in koreninske gnilobe.	bO l"e:z ni ras t_l"i:n' jE sE v"e: da v"EtS Od_n na p"a: da p E sn"i : _ dO li tS"i:Nk u S"i: mu S"i:ts in' p@r S"i:ts _ dO "@r jE in' kO rE n"i:n' skE gni l"O: bE	Obliko besede <i>plesni</i> naglasi kot glagol in ne samostalnik.
Pri Naurju je Kitajska ravnala podobno kot zdaj, saj ima tudi ta diplomatske odnose s Tajvanom.	pri na u:r ju jE ki t"a:j ska raU n"a: la pO d"o: bnO k"o:d zd"aj _ saj i m"a: tu di t"a: di p O m"a: tskE Od_n n"O: sE s t"a:j va nOm	<i>Kot</i> prepozna kot samostalnik <i>kót</i> in ne veznik <i>kot</i> . V pravilih se mora pri vezniku <i>kot</i> dopisati še možna vezava s časovnim prislovom.
Meni prinese sin Metod, s katerim živiva skupaj v Radovljici, največ češminovih vej iz doline Save v okolici Bodešč in Ribnega.	mE n"i : pri n"E: sE s"i:n' mE t"o:t s ka t"e: rim Zi v"i: va sk"u:paj w ra d"O:U lj tsi _ n"aj vEtS tSE Sm"i:nO vix v"E:j iz dO l"i:nE s"a: vE _ u O k"o: li tsi bO d"E:StS _ in' r"i: bnE ga	Zaradi težav s prepoznavanjem prilastkovega odvisnika prepozna besedo <i>meni</i> kot samostalnik <i>mení</i> in ne kot zaimek <i>mêni</i> .
Po drugi strani velja, da oblast, ki želi postopati preudarno, ve, da se mora samoomejevati.	pO dr"u: gi str"a: ni vE lj"a: d"a : Ob l"a:st ki ZE l"i: pO st"o: pa ti prE U d"a:r nO _ v"e: da sE m"o: ra s"a: mO O mE jE v"a: ti	Besedo <i>da</i> zopet naglasi zaradi težav s prepoznavanjem prilastkovega odvisnika. Besedi <i>da oblast</i> analizira kot samostojen stavek.

Tabela 2: Primeri napačnih slovničnih analiz

Slovnična analiza postavi v besedilo tudi stavčne oznake. Te sta bralca včasih pravilno intonirala, včasih pa sta na intonacijo pozabila. Lahko sta tudi pretiravala. Najmanj pa sta spremenila ton in način govora, čeprav bi se to pričakovalo pri velelniku. Označbe velelnosti bo treba verjetno še dodatno pregledati, ali so označene pravilno glede na bralčevo intonacijo.

Spodaj podajamo seznam vseh stavčnih oznak. Vse oznake so sestavljene iz dveh delov. Prvi določa vrsto in drugi naklon.¹⁷ Pri povedi [stan"u:jE u pEnzj"o:nu] tako označi, da gre za glavni stavek in povedni naklon. Na prvem delu oznake (STA:xy), torej na mestu x, so možne označbe:

- samostojni stavčni člen,
- glavni,
- vprašalni,
- pogojni,
- polstavek,
- osebkov,
- predmetni,
- prilastkov,
- prislovni.

Na mestu oznake y pa:

- povedni,
- velelni,
- pogojni,
- želelni,
- občudovalni in
- vezni.

Za še izboljšano analizo ali za podrobnejše jezikoslovne raziskave bo treba dodati še več oznak in transkripcijo podrobneje razdelati, a se je treba zavedati, da je pri našem delu cilj čim naravnejša sinteza, ki pa ne bo pretirano zapletena. S tega vidika je besedilo dobro označeno, Govorec 3 se bo lahko naučil in uporabljal različne intonacije. Najvažnejše bodo seveda vprašalna, velelna, povedna in pogojna. S stavčnimi oznakami se bo veliko lažje naučil tudi uporabljati primerno dolge premore pri različnih stavčnih oblikah in različnih stavčnih situacijah. Bolj kot razdelanost besedila je pomembna pravilnost vnesenih označb. To je bil tudi cilj naših popravkov v transkripciji.

¹⁷ Veliko oznak v transkripcijah ni enotnih s slovenistično stroko, saj so namenjene le sintetizatorju in so temu tudi prilagojene. Besedo naklon moramo tu razumeti kot intonacijo.

6.3 Grafemska pretvorba

Grafemska pretvorba je predelava besedila v računalniku razumljivo podobo. Treba je prepoznati vse znake (denarne enote ipd.), ki jih sintetizator nato tudi dobesečno prebere (znak @ pretvori in prebere kot [afna]). Številke pretvori v besede, prepoznati mora, ali gre za vrstilne ali glavne števnike, kratice – te lahko potem bere po slovarju izgovarjav skupaj ali pa črkovno, če so krajše od štirih črk (*CD* prebere kot [C"e:D"e:]). Skrajšane besede, kot je *g.*, pretvori v cele (*gospod*). Ločila mora prepoznati kot skladijska ali neskladijska, to prepozna po vnaprej določenih pravilih. Skladijska pika se npr. določi, če ji sledita presledek in velika začetnica, a le, če ji ne sledi številka ali če je na koncu odstavka in ji sledi prazna vrstica. Ločila (predvsem pika, klicaj in vprašaj) določajo tudi intonacijo in premore sintetiziranega glasu (a sintetizator to upošteva šele pozneje pri določanju prozodije). Samodejno se morajo prilagoditi obrazila iz pretvorjenega števnik v besedo, tako mora analizator poznati spol, sklon in število odnosnic. Z vsemi temi postopki nastane grafemski zapis, ki ga moramo nato pretvoriti v fonetičnega.

6.3.2 Grafemska pretvorba in Govorca 2 in 3

Stavčna analiza še ne ločuje med vezaji in pomišljaji, takšna ločila trenutno še prekinajo analizo, razen če se vezaj pojavi pri kratiki, pri kateri je dojet kot del besede. Te težave se pojavljajo, ker ločila še nimajo oznak MSD (oz. oblikoskladijskih oznak) v slovarju izgovarjav, saj so vsi ti grafemi določeni le kot ločila brez oblikoskladijskih vplivov. Pri več ločilih (npr. !!!, ?!?!?) prepozna samo prvo ločilo in naslednje obravnava kot ločene stavke. Prepozna krajevnih imen pa poteka le po slovarju izgovarjav in ne tudi po prepoznavi po velikih začetnicah, mestniških predlogih ali stavčnih funkcijah besed. Ne razpozna tudi premege govora, kar bo v prihodnje povzročalo še nekaj težav, saj se v sodobnejšem leposlovju pojavlja tudi označitev samogovora, dialoga s kurzivo in ne dosledno z narekovaji in dvopičjem. Tu se lahko seveda spet vrnemo k pomisleku o pomišljaju in vezaju. Njuna uporaba pri laičnih piscih je večinoma nepravilna. Težave z njima imajo v časopisih. Tudi v resnejših s precej naklade namesto pomišljaja uporabljajo vezaj. Natančna pravila o stičnosti in nestičnosti vezaja oz. pomišljaja bi tako lahko pri Besani (samodejnem

lektorju, ki deluje na istem sistemu kot Govorca 2 in 3 ter Presis) pri popravljanju povprečnega besedila pripeljala do več napak kot popravkov.

6.4 Transkripcija

Pri sintetizatorjih govora, ki so tema diplomskega dela, je pomembna predvsem točnost fonetičnega zapisa. Pravilno morajo biti zapisani tudi alofoni. Na elektrotehniški in računalniški konferenci v Portorožu leta 1995 so se pojavile začetne pobude za standardizacijo fonetičnega zapisa. Za fonetično zapisovanje je standardna uporaba abecede IPA (angl. International Phonetic Alphabet). Za preslikavo teh posebnih znakov uporabljamo mednarodno uveljavljene simbole MRPA, ki jim SAMPA služi kot osnova (Zemljak idr. 2002: 160). Osnovne značilnosti pisave MRPA so:

- širok glas se piše z velikimi črkami,
- ozek glas z majhnimi,
- naglas se označi pred soglasnikom naglašene samoglasnika z narekovajem ['],
- dolžina z dvopičjem [:],
- polglasnik z afno [@],
- glas črke *h* z znakom [x],
- mehčanost z znakom ['],
- dvoglasniški *u* se zapisuje z velikim *u* [U],
- *m* pred *f* ali *v* z [F],
- *n* pred *k*, *g*, *h* z [N],
- alofon fonema *j* z [I].

Vsi drugi glasovi, ki jih tu ne omenjamo, so zapisani že v Tabeli 1. V tabeli so tudi zapisi v mednarodni pisavi IPA. Transkribiranje pri Govorcu 3 smo z rabo in izkušnjami prilagajali. „Enako kot vsa načela glede gradnje korpusa so tudi načela za transkribiranje korpusa odvisna od namembnosti korpusa /.../ Fonetik potrebuje manjšo količino podatkov, vendar morajo biti ti mnogo boljše detajlno transkribirani v smislu prozodije in akustične

realizacije ter nujno povezani z zvočnimi posnetki“ (Zemljarič Miklavčič 2008: 93). Pri raziskovanju drugih govornih zbirk se srečujemo predvsem s tistimi, ki so namenjene raziskovanju spontanega govornega jezika. V primeru Govorca 3 pa „/b/ranih besedil ne razumemo kot pravi govorni jezik, saj gre samo za oralizacijo pisnega jezika“ (Zemljarič Miklavčič 2008: 23). Gorjanc korpuse, ki so sestavljeni iz vnaprej zbranih besedil in so studijsko posneti, imenuje korpusi govora ali govorne zbirke. Drugačni so referenčni korpusi, pri katerih jezikoslovci zanimajo predvsem skladenjski vzorci in pomeni besed pri spontanem govoru. Pri teh se ponavadi uporablja ortografska transkripcija. „Čeprav se načrtovalci vsakega korpusa odločijo za svojo transkripcijsko shemo glede na namen korpusa in na specifične lastnosti jezika ter jezikovne situacije, je gotovo smiselno preučevati obstoječe transkripcijske modele in se učiti na njih“ (Zemljarič Miklavčič 2008: 108).

Obstaja več vrst transkripcij, ki se razlikujejo po zvestobi do akustične realnosti posnetega govornega signala. „Otrografska transkripcija ali transliteracija za zapisovanje govornih besedil uporablja standardni (knjižni) zapis besed. Fonemska transkripcija temelji na fonemih izbranega jezika, alofonska transkripcija uporablja različne oznake za foneme v različnih okoljih, fonetična transkripcija pa zapisuje individualno izgovarjavo besed posameznega govornika“ (Zemljarič Miklavčič 2008: 105). Poleg fonetičnega zapisa se uporabljajo tudi nekatere prozodične oznake. TEI (priporočila za označevanje govornih besedil) priporoča na primer premor, neverbalne glasove, kinezične dogodke (kimanje, odkimavanje), pisanje ... Tako „/s/topnja uvajanja fonetičnih oznak v transkripcije niha od nič do zelo natančne fonetične transkripcije“ (Johansson 1995: 93 v Zemljarič Miklavčič 2008: 100).

V Zemljarič Miklavčič (2008) najdemo stopnje označevanja govornih zbirk. Te so:

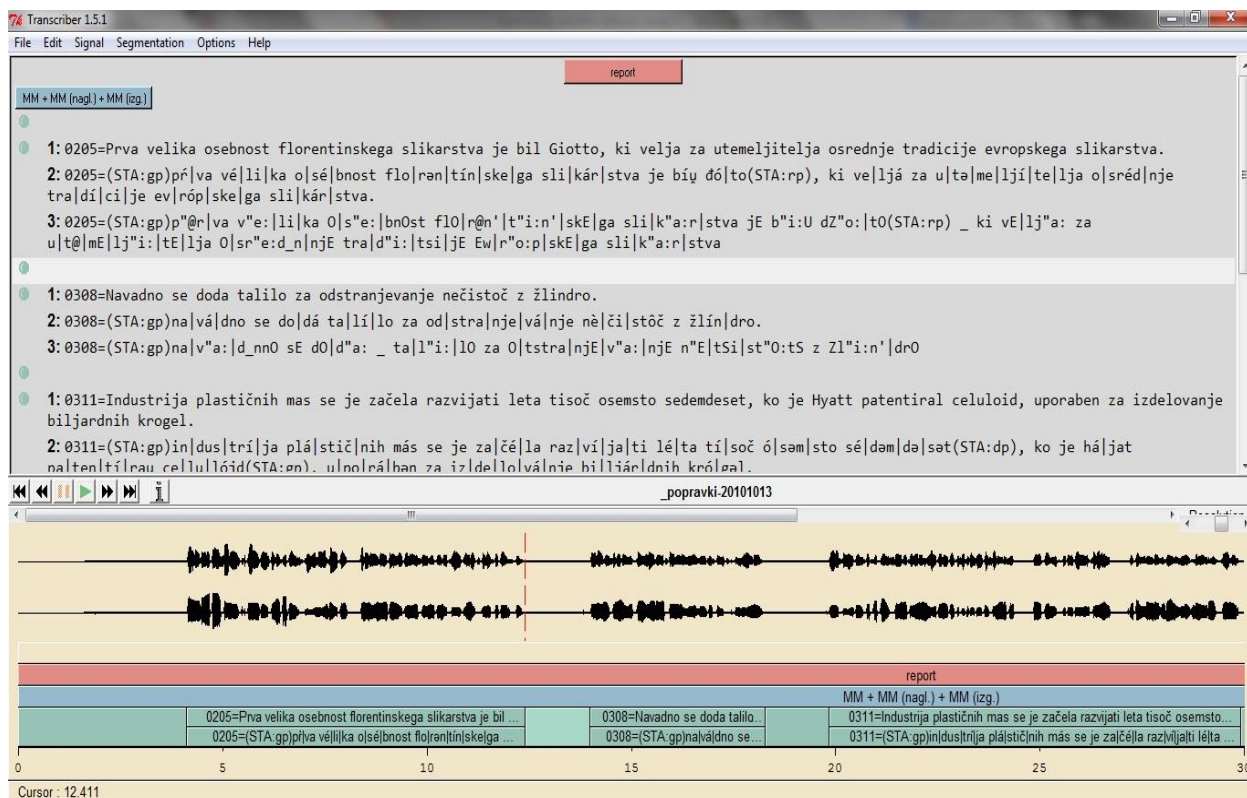
- ortografska raven (standardni zapis),
- citatno-fonemska raven (zapis s fonemi, zapis iz zapisa besede v izolaciji),
- širša fonetična raven (upošteva se izgovor besed v kontekstu),
- ožja fonetična raven (natančen zapis govora z variantami fonemov),
- akustičnofonetična raven (ločuje vse zaznavne elemente na spektogramu).

6.4.1 Transkribiranje in Govorec 3

Pri transkribiranju se srečamo z mnogimi izzivi. To je najpomembnejši del za jezikoslovca, ki sodeluje pri ustvarjanju umetnega govora, še posebej ko gre za vrste sintetizatorjev, kot je Govorec 3, ki ima po Zemljarič Miklavčič četrto stopnjo označevanja (ožja fonetična raven).

Transkripcije smo zapisovali v programu Transcriber, ki se od drugih najbolj izkaže pri veliki zbirki posnetkov in transkripcij.¹⁸ V spodnjem delu zaslona je tudi grafični prikaz govornega in laringografskega signala. Tako smo lahko lažje ločili med zvenečimi in nezvенеčimi glasovi ter odmori, in sicer brez natančnejše sonogramske analize. Ima enostaven uporabniški vmesnik, večina ukazov je dosegljiva s tipkovnico, popravki mej posameznega posnetka se nastavijo kar s klikom na srednji gumb miške in z njenim premikom. Lahko se predvaja le en kanal stereoposnetka. Ker je bil na desnem kanalu posnet laringografski signal, smo poslušali samo levi kanal v mononačinu. Edine težave so povzročali nenamerni pritiski na vračalko, ker so povedi potem razdelili tudi na govornem signalu. Transcriber je brezplačen in je objavljen pod licenco GNU (angl. General Public License). Njegov avtor je Claude Barras, prva verzija je bila registrirana leta 2001. Posnetki so po TEI segmentirani na izjave, pri Govorcu je to vsaka posamezna poved, ki je tudi oštevilčena. Vsaka izjava je zapisana v treh pisavah: v navadni z vsemi ločili in pravopisnimi pravili, v drugi vrstici so že postavljeni naglasi, slovnična analiza in odmori, v tretji pa je fonetični zapis v pisavi MRPA. Druga vrstica je pomembna predvsem za primerjavo s slovarjem izgovorjav. Ko bodo vneseni vsi popravki, se bodo naglasi, izgovori polglasnikov in drugih glasov primerjali s tistimi, ki so zapisani v samem slovarju. Velika pomanjkljivost Transcriberja je nezmožnost akustične analize, ki bi v nekaterih primerih olajšala in omogočila natančnejšo transkripcijo. Slika 8 prikazuje delovno okolje za popravljanje Govorčevih transkripcij.

¹⁸ Več o različnih transkripcijskih orodjih in njihovih prednostih in slabostih v monografiji *Govorni korpusi* (121–128).



Slika 9: Delovno okolje v Transcriberju

Pri Govorcu 3 so bile fonetične transkripcije samodejno tvorjene iz grafemske transkripcije po pravilih, ki so jih sestavili na Amebisu, d. o. o., po priporočilih Instituta „Jožef Stefan“. Ta pravila smo že precej popravili in dopolnili, saj so vsebovala nekaj napak. S tem smo se tudi izognili večjemu in obsežnejšemu poseganju v slovar izgovorjav. Pri preverjanju rezultatov se je hitro ugotovilo, da je lahko teorija v jeziku (predvsem govornem) precej relativna in nezanesljiva. V začetnih fonetičnih transkripcijah so se uporabljale vse variante fonemov, upoštevali so se premene po zvonečnosti in vplivi sosednjih besed na sam izgovor konca in začetka besede. Te vplive (premene po zvonečnosti) na medbesedni ravni je bilo treba velikokrat popravljati, saj sta bralca delala med besedami krajše neslišne premore in se tako oddaljevala od teorije. Vsi premori se v transkripcijah označujejo s tremi znaki, s podčrtajem [_] čiste tišine, z znakom [<] vzdih in z [>] izdih. Dobro bi bilo različno označevati krajše in daljše premore čiste tišine, kot so na primer označeni v korpusu London-Lund (z dolгим in kratkim podčrtajem). Težave smo imeli pri označevanju odmorov z različnimi oznakami, saj so se lahko med tišinami pojavljali cmoki in drugi nenamerni zvoki bralcev ali pa se je hkrati pojavila kombinacija več govornih pojavov (npr. tišina, vdih in cmok). Predvsem smo gledali na grafični prikaz govornega signala, na

katerem se je videlo, ali je med besedama daljša tišina ali dih. Z odmori so se na Amebisu, d. o. o., že ukvarjali, a smo pri delu s transkripcijami govornih zbirk naredili še nekaj sprememb. Odmore sta napovedovalca uporabljala precej različno. Moški napovedovalec ima krajše in samo stavčno strukturo bolj kot z odmori izraža z intonacijo, napovedovalka pa uporablja daljše in jasnejše tišine med stavki.

Polglasnik se je pojavljal veliko pogosteje, kot smo pričakovali in kot so jih pri grafemsko-fonemski pretvorbi predvidevala pravila. Za natančnejšo transkripcijo bi potrebovali akustično analizo, saj je pri hitrem govoru težko slišno razločiti, ali gre res za polglasnik. Bralca sta ga vedno izgovarjala, če je bil široki nenaglašeni *e* na koncu besede v zaprtem zlogu. Navajamo nekaj primerov: [pOtrEb"u:j@S], [pos"e:b@j], [m"ajSp@rk], [d"O:b@r], [v"e:t@r], [s"i:ts@r], [m"E:nEdZ@r], [x"i:t@r] ... Pojavljala so se tudi neskladja s samodejno pretvorbo in govorom bralcev pri nenaglašanih besedah, nekaj primerov smo že navedli. Za primer analize smo vzeli členek *že*. Pri popravljanju transkripcije smo si sproti zapisovali, kolikokrat kdo od napovedovalcev izreče *že* s širokim in kolikokrat z ozkim *e*. Pri štetju nismo izključevali nenaglašanih ali naglašanih, obravnavali smo jih enako.

Moški napovedovalec		Ženska napovedovalka	
[Ze]	[ZE]	[Ze]	[ZE]
11	18	19	10

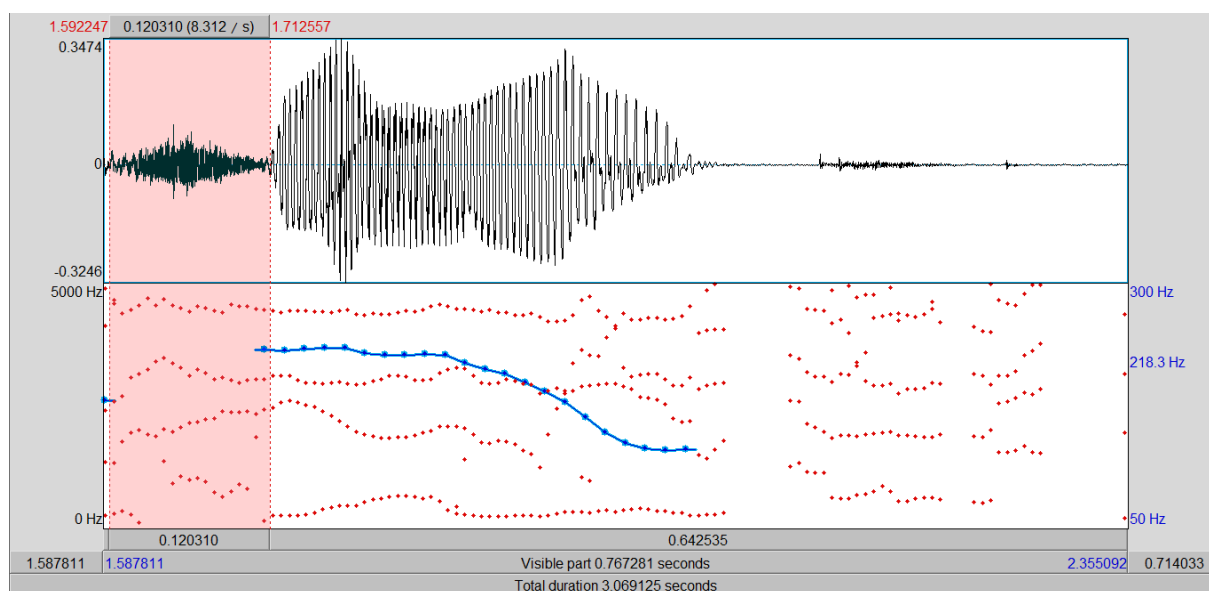
Tabela 3: Primerjava izgovora členka *že* pri bralcih

Bralca bereta členek zelo različno, na izgovor pa glasovna okolica ne vpliva. Pri bralcih nenaglašenost ne pomeni nujno širine glasu. Naglašenost ni vključena v analizo, saj v programu Transcriber akustična analiza ni mogoča.

Precej dilem je povzročal tudi glas *j*, ki ga pišemo z *i* in sledi samoglasniku (npr. *celuloid*, *pacient*). Po predlogu Instituta „Jožef Stefan“ se je najprej posebej označeval z [I], a smo ta grafem oz. glas ukinili. Namesto tega smo glas zapisovali tako, kot sta ga napovedovalca dejansko izgovarjala, z [i], [ij] ali [j]. Tako smo zmanjšali število glasov in olajšali analizo. Te težave pa bo treba premostiti še pri samodejni pretvorbi grafemskega

zapisa v fonetičnega. Predlagamo, da se prejšnji [I] v pravilih pri besedah, kot je *pacient*, nadomesti z glasovoma [ij] ali [i] in pri besedi tipa *celuloid* samo z [j].¹⁹

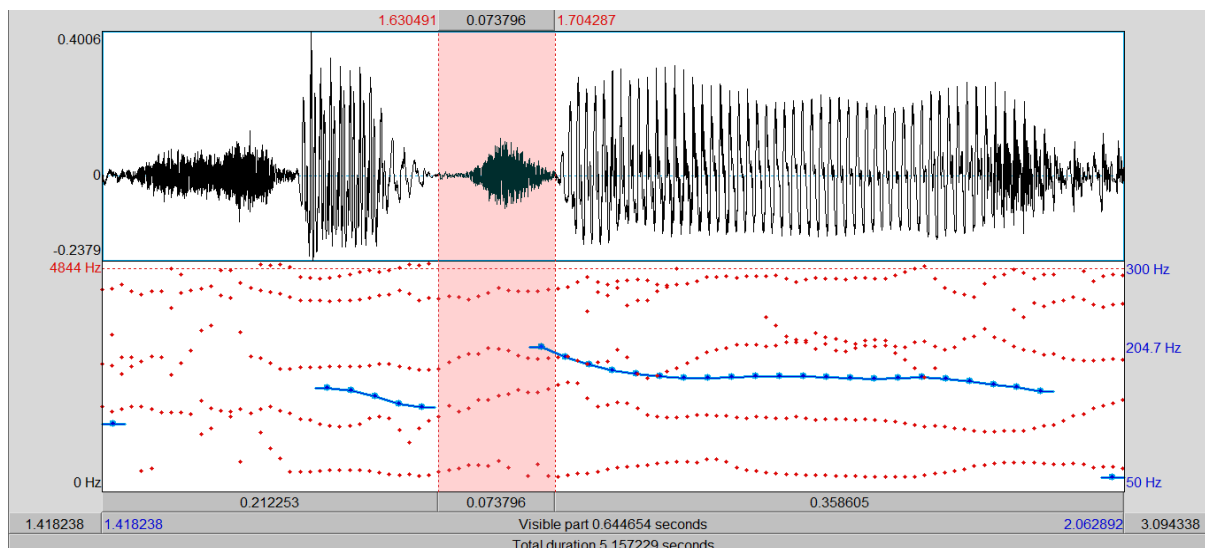
Opravili smo tudi nekaj akustičnih analiz. Sicer je testnih vzorcev premalo, da bi bili reprezentativni za celotno govorno zbirko, a rezultati kažejo, kje bo za boljše delovanje Govorca potrebna podrobnejša analiza. V primerih, v katerih se pričakuje izgovor [ij], je velikokrat prišlo do redukcije glasu *i* ali pa se *i* sicer sliši, a se pri akustični analizi pokaže, da nima osnovnega tona in je ustvarjen le s turbulencami v ustni votlini brez pomoči glasilk. Ali gre v tem primeru za en glas, ki ga slišimo kot dva [tsi], ali za dva glasova ali pa za kakšen poseben nezveneči *i*, je treba še posebej raziskati. Na Sliki 10 vidimo analizo besede *cianid*, ki jo je prebrala bralka, označen pa je glas [tsi]. Modra črta označuje osnovni ton, tako ugotovimo, da ta glas ni zveneč.



Slika 10: Analiza besede *cianid* (bralka)

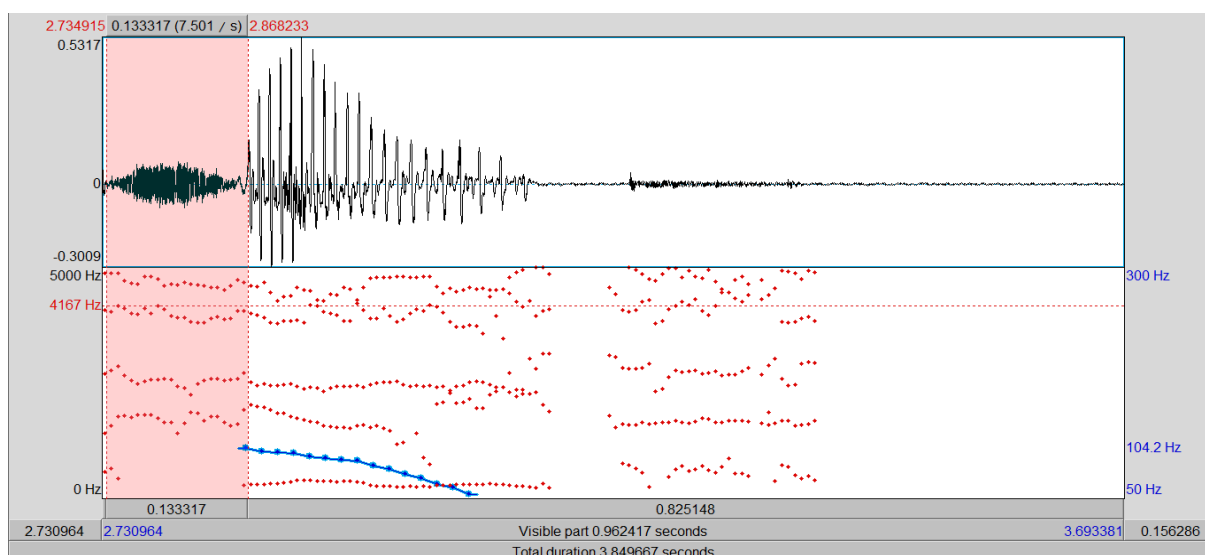
Enako je tudi pri podrednem delu besede *socialnoekonomski*. Pri analizi smo analizirali samo prvi del besede *socialno-*. Spet se pojavi glas [tsi], ki je sicer manj izrazit, *i* se sliši le, če ga poslušamo izolirano. Gre za isto bralko kot v prejšnjem primeru.

¹⁹ Govorec je namenjen predvsem slabovidnim osebam, zato smo pri delu velikokrat sledili njihovim potrebam. Predlagamo tudi, da sintetizator besede, kot je *cianid*, bere tako, kot so zapisane, in ne po SP 2001, ki narekuje izgovor [tsijan"i:t], saj je za slabovidne izredno pomembno, da slišijo besedilo tako, kot je zapisano.

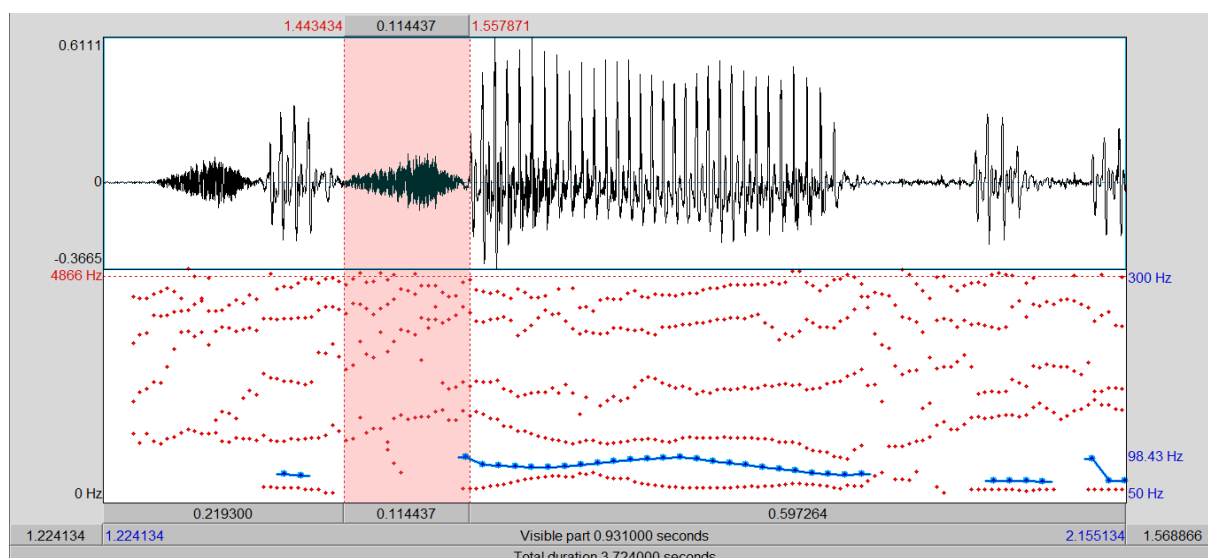


Slika 11: Analiza besede *socialnoekonomski* (bralca)

Tudi pri bralcu se pojavi isti glas pri istih besedah. Prilagamo dva primera, in sicer *cianid* in *socialno*-. Tako lahko trdimo, da pojav ni povezan s posameznim bralcem. Ali je ta pojav posledica istega šolanja govora oz. pravorečja, bi morali preveriti na več in različnejših bralcih.



Slika 12: Analiza besede *cianid* (bralec)



Slika 13: Analiza besede *socialnoekonomski* (braliec)

V Tabeli 4 podajamo primere izgovora glasu *i* pred samoglasnikom. Primeri glasu [tsi], ki smo jih akustično analizirali, so kot posebnosti zapisani krepko. V teh primerih je lahko *i* bolj ali manj reduciran ali pa bolj ali manj izraziteje zrasel z glasom *c*.

Beseda	Izgovor bralcev v bazi []
kvocient	kv"o:tsijEnt
hieroglifi	xjErOgl"ifi
ioni	ij"o:ni
koeficienti	kOEfitsij"E:nti
postsocialistični	p"o:stsOtsjal"i:stitSni
genialna	gEnij"a:lna
variira	varij"i:ra
cianid	tsjan"i:t
radio	r"a:dijO
ambiciozen	ambitsj"o:z@n
fizioterapevt	f"i:zijOtErap"e:Ut
tiare	tij"a:Re
ciankalij	tsjaNk"a:lij
komediograf	kOmEdiOgr"a:f
heliosa	x"e:liOsa
niansa	nij"a:nsa
pionirska	piOn"i:rska
nacionalizma	natsjOnal"i:zma
aksiom	aksj"o:m
varianta	varj"a:nta

Tabela 4: Primeri izgovora *i* pred samoglasnikom

Poleg pojavov, ki smo jih našli, je treba v zapisih transkripcij rešiti še glasovni pojav, ko si v glasovni verigi sledita dva soglasnika in lahko pride do zlitja v enega dolgega. Za zdaj, dokler še ni znanih rezultatov samodejne analize, smo v transkripciji pustili te dvojne grafeme zapisane, kot so. Beseda *najoddaljenejši* je tako v transkripciji zapisana kot [n"ajOdd"a:lJEnEjSi]. Pri obeh bralcih smo izmerili nekaj primerov trajanja zlitih in samostojnih glasov (Tabela 5). Ti so se pojavljali v govorni zbirki, zadnji pa je primer, ko gre za tri enake glasove enega za drugim. Tudi v tem primeru bi bilo dobro izmeriti še, kaj se zgodi z dolžino. Žal bralka vseh treh glasov ni prebrala zlito. Besedilo je prebrala kot [um"e:s z zOb_nm"i:], bralec pa je združil vse tri glasove v enega in tako prebral kot [um"e:z z zOb_nm"i:]. Dolžina vseh treh je bila 0,167 sekunde. V primerjavi z naslednjo analizo vidimo, da je tri zlite glasove izgovoril za približno tretjino daljše kot dva.

Primeri
v uradu
s stigmo
razstavi
čas za
razsaja
v vlogo
medtem
izzvali
vmes z zobmi

Tabela 5: Primeri besed, v katerih si v verigi sledita dva soglasnika

Rezultati krajše analize bralkinih dolžin so v Tabeli 6. V oklepaju je število primerov, s katerimi smo analizirali glas in izračunali povprečje. Meritve smo opravili za glasove *s*, *z*, in *d*. Dolžina je podana v sekundah.

Glasovi	s	z	d
zliti	0,119 (1)	0,140 (1)	0,075 (1)
enojni	0,115 (3)	0,070 (3)	0,036 (2)

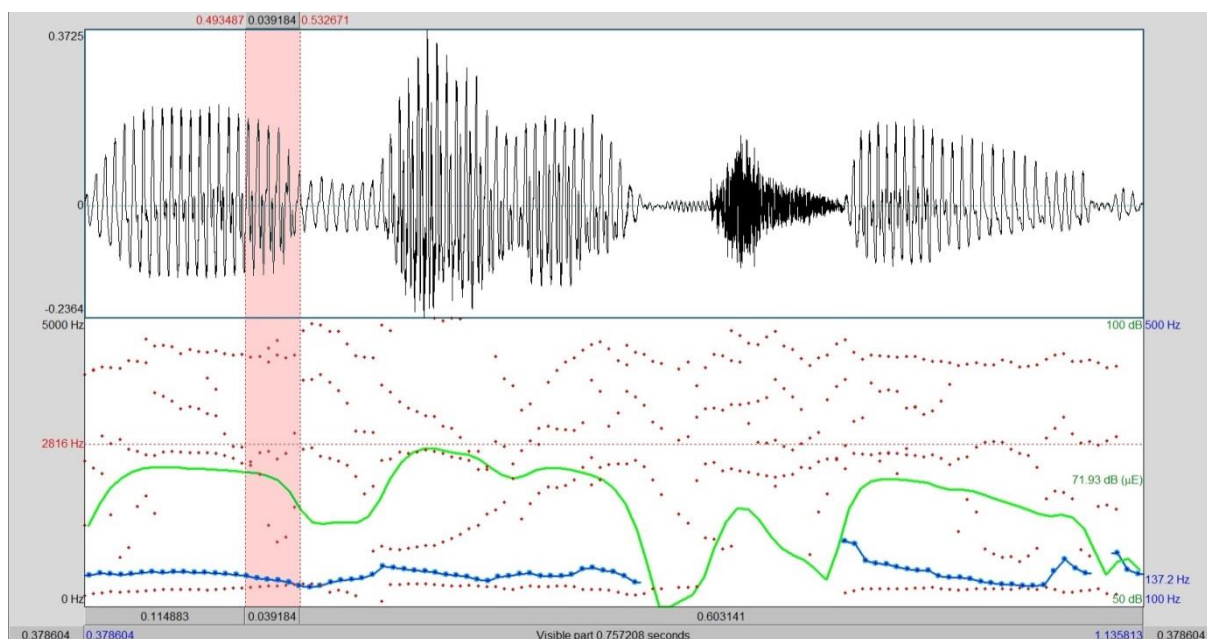
Tabela 6: Časovna analiza zlitih glasov (bralka)

Zlita dvojna glasova *d* in *z* sta približno dvakrat daljša kot enojni glas, zato bi bila zanju potrebna podrobnejša analiza na večjem vzorcu, da bi se domneve potrdile. Medtem ko pri glasu *s* ni zaznati razlik. Pri bralcu navajamo rezultate iste analize v Tabeli 7. V oklepaju je poleg rezultatov zopet zapisano število primerov, na katerih je izračunano povprečje.

Glasovi	<i>s</i>	<i>z</i>	<i>d</i>
zliti	0,140 (2)	0,132 (3)	0,108 (1)
enojni	0,121 (3)	0,073 (2)	0,052 (3)

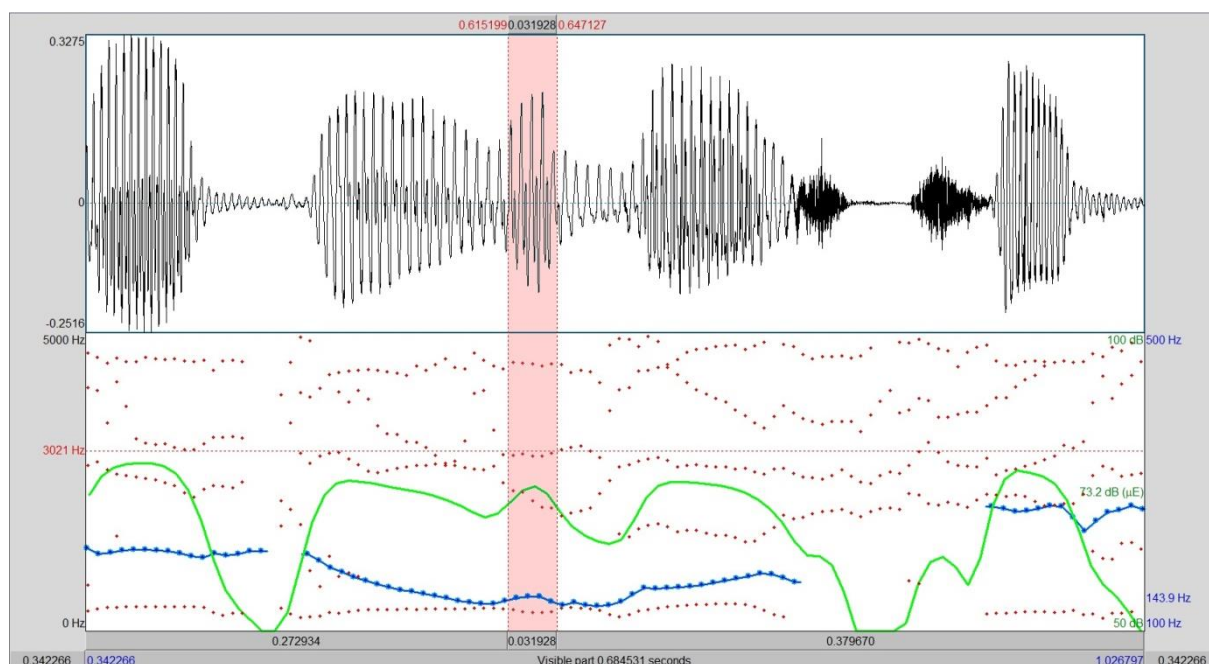
Tabela 7: Časovna analiza zlitih glasov (bralec)

Rezultati so podobni kot pri bralki. Pri *z* in *d* je zliti glas približno dvakrat daljši, medtem ko lahko pri glasu *s* govorimo le o 14-odstotni razliki, kar je posledica nižjega povprečja dolžine le enega, izjemoma krajšega izmerjenega glasu. Ker je pri obeh bralcih rezultat podoben, bi bilo dobro o zlitih soglasnikih narediti obširnejšo raziskavo in ugotoviti, ali je res, da so zliti glasovi približno dvakrat daljši in da je pri tem glas *s* izjema. Tako kot smo naredili krajšo analizo zlitih glasov *s*, *z* in *d*, smo pregledali tudi zvoneči *y*, ki smo ga v transkripciji označevali kot [G]. Prva bralkina besedna zveza, analizirana v Praatu, je *porodnih bolečin*. Na Sliki 14 je analiza dela [nix bOlEtS"i:n']. Glas [G] je označen z rdečim senčenjem.



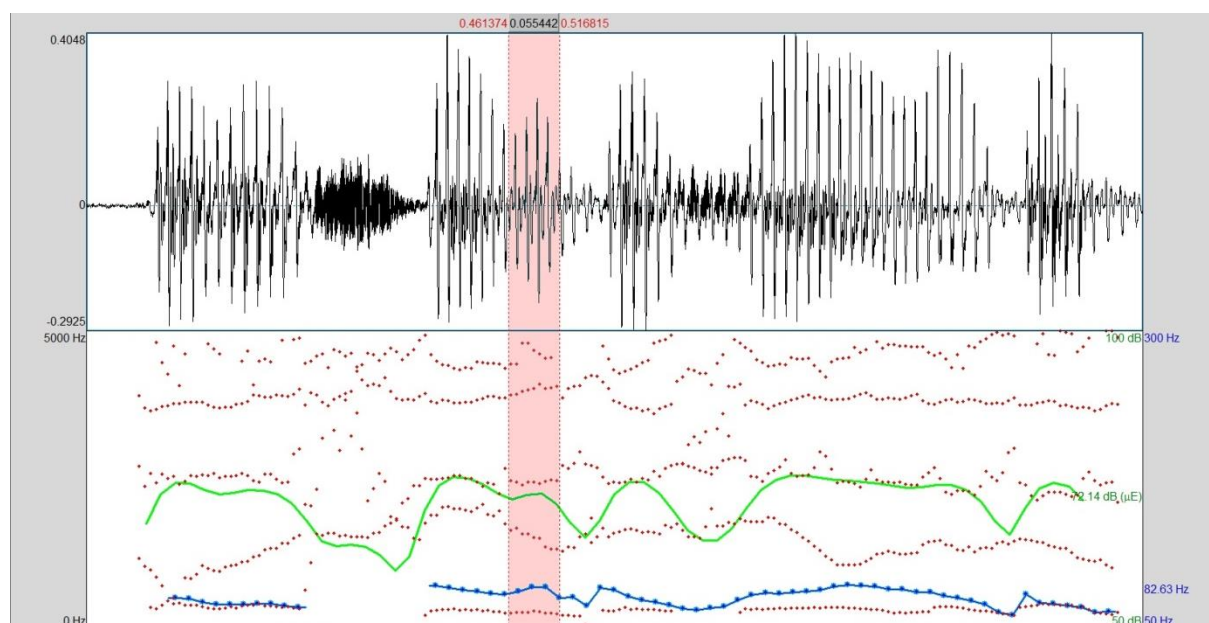
Slika 14: Analiza zvonečega *y* (bralka)

Slika 15 prikazuje besedno zvezo *parketnih deščic*, in sicer del [niG d"@StSits].



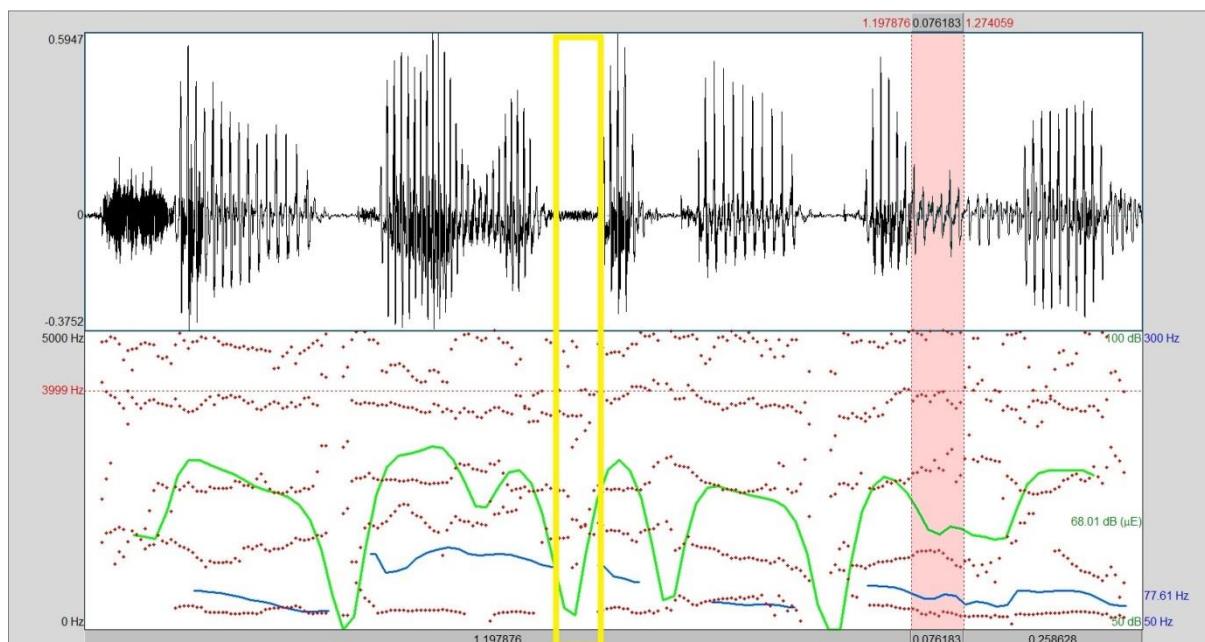
Slika 15: Analiza zvonečega γ (bralka)

V obeh primerih vidimo, da gre res za zvoneči γ, saj ima glas osnovni ton, ki ga dokazuje modra črta v analiznem oknu. Te glasove smo našli tudi pri bralcu. Na Sliki 16 je primer v *vaseh deževnega gozda*. Zvoneči γ je zopet označen z rdečim senčenjem.



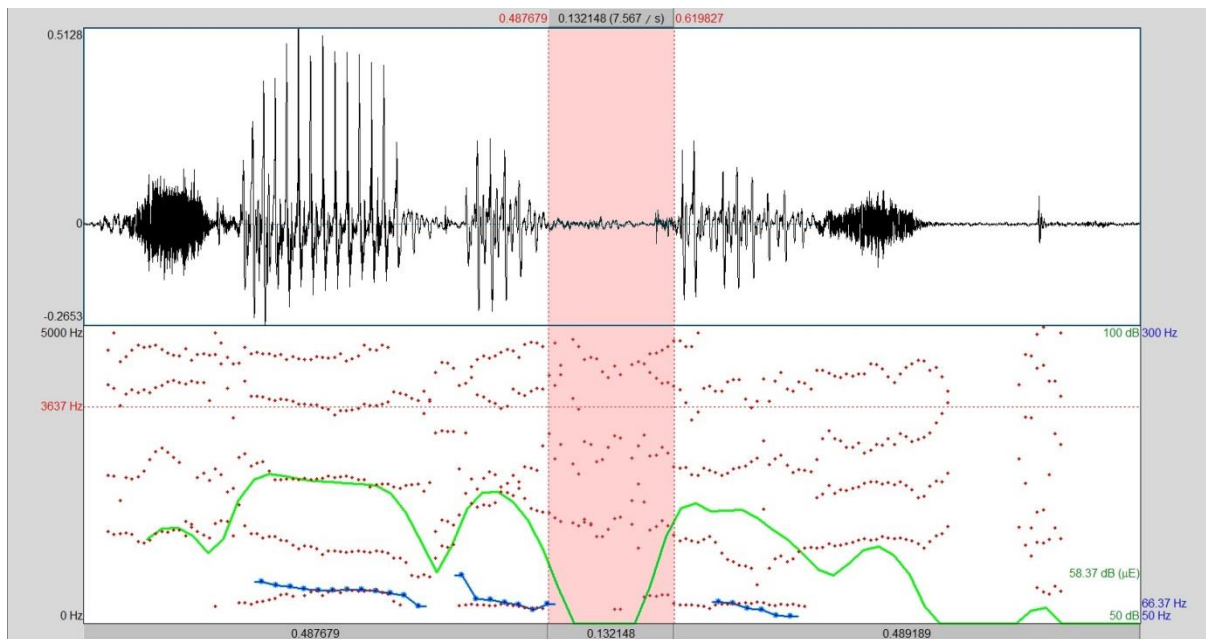
Slika 16: Analiza zvonečega γ (bralec)

Dodajamo še analizo izseka [s@m na t'e:Zjix Et'a:paG b'i:U]. Poleg rdečega osenčenega glasu [G] je označen še nezveneči [x]. Tu se dobro vidi, da prvi glas x nima osnovnega tona, drugi y pa ga ima. Gre torej za dve varianti glasu x.



Slika 17: Analiza zvonečega y s primerom nezvonečega x (bralec)

Grafemsko-fonemska pretvorba je zvezo *slabih družb* pretvorila v zapis [sl'a:biG dr'u:Zp]. Ker še ni pravila, ki bi upošteval dva glasova nazaj za premeno po zvonečnosti, smo morali najprej popraviti besedo [dr'u:Zp] v [dr'u:Sp]. Nato smo v zapisu popravili zvoneči y v nezvonečega, saj smo lahko že v Transcriberju pri posnetku glasilk videli, da je kljub pričakovani zvonečnosti bralec x izgovoril brez pomoči glasilk, in sicer nezvoneče. Ko pa smo ta del posnetka še podrobneje akustično analizirali v Praatu, se je pokazalo, da je prišlo do še ene izjeme – namesto glasu [d] prebere bralec [t]. Zapis je bil na koncu: [sl'a:bix tr'u:Sp]. Na posnetku (Slika 18) sta označena glasova [x] in [t], nezvonečnost sicer traja še nekje do polovice glasu [r], a se bralcu nato vklopijo glasilke. Pri tem primeru lahko dobro vidimo, kako pravila ne morejo predvideti vsega in da je človeški glas precej nepredvidljiv. Za natančnejše rezultate bi potrebovali podrobno akustično analizo, ki odkrije marsikaj, kar je ostalo ljudem neslišno in prek grafičnih zapisov govornih signalov nevidno.



Slika 18: Analiza primera [sl'a:bix tr'u:Sp]

S temi akustičnimi analizami smo nakazali nekaj praktičnih izkušenj, s katerimi smo se srečevali pri popravljanju transkripcij. Analize so sicer kratke, a jasno nakazujejo, kaj vse bo še treba postoriti za izboljšavo sintetizatorja. Nekaj teh je:

- podrobnejši pregled polglasnikov,
- analiza dolžine zlitih soglasnikov,
- podrobnejši pregled variant glasov (primer sta zveneči γ oz. [G] in glas [tsi]),
- vnos novih pravil (npr. premena po zvonečnosti za več kot en glas nazaj).

7 Anketa in rezultati

Anketo, ki je dodana kot priloga, smo razdelili šestim slovenistom. Poleg vprašalnika smo posredovali tudi transkripciji in posnetke dveh besedil, ki smo jih sestavili in sintetizirali s pomočjo Govorca 2. Besedili sta sintezi dveh bralcev, in sicer Renata in Mateja. Vsako besedilo vsakega bralca smo poslali še v treh različnih bralnih hitrostih (počasna, normalna in hitrejša). Tako so anketiranci poleg osnovnega vprašalnika s transkripcijama dobili še 12 posnetkov sintetiziranega glasu. Besedili smo sestavili tako, da sta vsebovali pomembnejše govorne posebnosti, ki smo jih obravnavali pri transkripcijah Govorca 3. Prvo besedilo je klasično javno besedilo v zborni slovenščini, drugo pa je skupek povedi v slengu in narečju, ki bi se pojavljale v zasebnih govornih položajih.

	Ocenjevalci						
Ocena	1	2	3	4	5	6	Povprečje
glasovna ustreznost	3	3	5	3	5	2	3,5
naglasna ustreznost	3	5	5	3	4	2	3,7
barva glasu	4	4	5	2	5	4	4
intonacija	4	2	1	1	4	3	2,5
register	4	3	1	2	5	3	3
poudarki	2	2	2	1	5	2	2,3
dolžina in mesto premorov	2	4	3	1	4	2	2,7
uporabnost uravnavanja hitrosti govora	5	4	5	5	5	4	4,7

Tabela 8: Številčni rezultati ankete

Na začetku so anketiranci s števili od 1 do 5 ocenjevali kakovost govora. V Tabeli 8 so izpisane ocene in povprečja vseh ocenjevalcev. Najboljše so ocenjeni glasovna in naglasna ustreznost, barva glasu in uporabnost nastavljanja hitrosti govora. Težave v kakovosti se pričakovano pojavijo pri intonaciji in registru glasu, premorih in poudarkih (premore so tudi v transkripcijah pri Govorcu 3 povzročali največ preglavic). Pri podrobnejših vprašanjih so

anketiranci kot najbolj moteče in najmanj kakovostne elemente navajali predvsem intonacijo in uporabo premorov. Motila sta jih še izgovor glasu v v različnih glasovnih položajih (pred soglasniki in kot dvoglasniški μ) in naglaševanje zloženk ([d"E:lOdaj"alts@U]). Kot težavne so omenjali tudi barvo, register in hripavost glasu. Samo tretjina anketirancev bi Govorca 2 uporabljala kot bralnik ali kot del aplikacij tehničnih pripomočkov.

Rezultati ankete so pričakovani, vzrok nekaterih motečih elementov je predvsem v naravi delovanja difonskega sintetizatorja, ki lahko dvoglasniški μ zaradi fonetičnih pravil (preverjena bolj v teoriji kot v praksi) nenaravno vsiljuje v jezik, to se bo z Govorcem 3 izboljšalo, saj je primerov izgovora glasu v v bazi veliko več in so izgovorjeni v najrazličnejših glasovnih okoljih, natančneje pa smo uredili tudi pravila grafemsko-fonemske pretvorbe.

8 Zaključek

Sintetizatorji govora imajo v slovenščini še veliko možnosti za razvoj in napredek, predvsem bi se morali v raziskave vključiti tudi slovenisti. Medtem ko sta slovnica in pravopis že dodelana, pa je govorjeni (standardni) knjižni jezik še v večini nedoločen in neraziskan. Seznam glasov v Slovenskem pravopisu je teoretično dodelan, a prihaja v praksi do razhajanj. Novejše ugotovitve o zvenečih in nezvенеčih ustničnoustničnih šumnih glasovih ([w] in [W]) so premalo poznane, tako so avtomatika in starejša grafemsko-fonemska pravila in predvsem zanašanje neslovenistov na razčiščeno normo v Slovenskem pravopisu povzročili v transkripcijah nekaj težav, ki jih bo treba ročno popraviti, saj bralca v pred glasovoma *l* in *r* bereta večinoma kot zobnoustnični *v*. Glede na zadnje rezultate raziskav med jezikoslovci smo pri Amebisu, d. o. o., ustničnoustnični šumni (zvenci in nezvenci) *v* ukinili, to smo izpeljali tudi že v transkripcijah.²⁰ Težava ni, da uporabniki slovenistične stroke ne bi upoštevali, pač pa da slovenisti rahlo zamujamo in zaostajamo za potrebami hitro spreminjajočega se sveta. Analize, kot smo se jih lotili v diplomskem delu, niso namenjene le slovenistični strokovni javnosti, njihova dodana vrednost je, da so posredno uporabne v informatiki, poslovnem svetu in zabavni industriji. Takšni sintetizatorji, kot je Govorec, se lahko uporabljajo v mobilnih aplikacijah, e-bralnikih in e-asistentih, s takšnimi pripomočki se lahko zmanjšajo stroški poslovanja podjetij, olajša vsakdan slabovidnim in obogati izkušnja z zabavnimi aplikacijami. Glede na to, da so tudi slovenska narečja temelj humora, bi bili razvoj mobilne aplikacije z osnovami narečnosti ali posnetki narečnih govorcev za uporabnike izredno zanimivi in lahko tudi dobičkonosni, poleg tega pa še seveda priložnost slovenistične stroke, da javnosti približa jezik in znanje, ki ga premore in je že zapisano ter v večini raziskano.

V diplomskem delu smo na kratko povzeli zgodovino in osnovne lastnosti sintetizatorjev govora. Prvi mehanični sintetizatorji so se pojavili že v drugi polovici 18. stoletja, pravi sintetizator govora z imenom VODER pa leta 1930. V šestdesetih letih so razvili formantne sintetizatorje, ki so obsegali do 100 enot v besedišču, število pa je že v sedemdesetih letih zraslo do 1000. Potem se je že začel razvoj sintetizatorjev, ki temeljijo na lepljenju enot (metoda LPC), ta metoda je tudi temelj Govorca 2. Nazadnje so se pojavili še

²⁰ Več v člankih Hotimirja Tivadarja *Fonem /v/ v slovenskem govorjenem knjižnem jeziku* (1999) in *Še o fonemu /v/ in njegovih alofonih* (2000) avtorice Tatjane Srebot-Rejec.

sintetizatorji, ki temeljijo na metodi PMM, in sinusni sintetizatorji, varianta teh pa je Govorec 3. Pojasnili smo njihov način delovanja in vse potrebne elemente, ki so nujni za uspešno sintezo (govorna zbirka, slovnična analiza, transkripcije in grafemsko-fonemska pretvorba, slovar izgovarjav ter dobra sinteza glasu). Podrobneje smo opisali svoje izkušnje z delom pri transkripciji in naglasih besed, z akustično analizo pa pokazali na nekaj pojavov, ki potrebujejo konkretnjšo in obsežnejšo analizo. Prihodnost jezikoslovja (predvsem slovenščine kot jezika z manjšim številom govorcev) je v digitalni humanistiki. A bodo za uspešno izpeljavo takšnih projektov, kot je Govorec, zopet potrebne nekatere osnovne raziskave jezika (tako fonetične kot oblikoslovne) z novimi temelji in novimi cilji glede na potrebe informacijskih tehnologij. Anketa je pokazala, da je potreben novejši sintetizator z naravnejšim govorom, to je seveda tudi cilj dela sodelavcev na Amebisu, d. o. o., in moje jezikoslovne pomoči pri Govorcu 3.

9 Povzetek

Pri delu z Govorcem 3 in raziskovanju za diplomsko nalogo smo opazili, da je v slovenskem jeziku še veliko nedorečenega, da bo treba opraviti nove fonetične raziskave, dozdajšnja spoznanja, ki se razlikujejo od SP 2001, pa na novo normirati. Vprašljivo je uvajanje obvezne širine nenaglašanih glasov *e* in *o*, pojavljajo se težave pri ozkem izgovoru naglašanih samoglasnikov v tujkah, izgovori s širino, ki so pogosti, tako odstopajo od norme. Neenotnost se pojavlja v SP 2001 tudi pri naglaševanju (npr. dvánajst in dvanájst), predvsem zaradi nedoločenosti osnove za knjižni jezik.

Slovenistična stroka se premalo posveča digitalni humanistiki in s tem tudi sintezi govora. Informatiki tako sami rešujejo vprašanja, ki bi jih morali reševati jezikoslovci, popravki napak in izboljšanja jezikovnih tehnologij so zato dolgotrajnejši, razvoj pa se upočasnjuje. V jeziku bi bilo treba najti sistem, ki bi ga lahko razumel tudi računalnik. Tako smo z določenimi novimi grafemsko-fonetičnimi pravili na Amebisu rešili veliko težav pri prevajanju besedila v fonetični zapis. Dobro nam je uspelo določiti na primer pravila za izgovor črke *v*, zanjo nam ni treba več vnašati posebnih fonetičnih zapisov v slovar izgovorjav. V zadnjem mesecu smo v transkripcijah ukinili večino alofonov, saj ti z naravo delovanja metode HNM niso več potrebni. Fonetična pravila so nujna in morajo biti jasna, a je treba izpostaviti, da naravni govor od teh zakonitosti velikokrat odstopa. V naši govorni zbirki smo našli tudi nezveneči glas *i*, ki do zdaj ni bil poznan, pojav bo treba podrobneje raziskati. Raziskave so bile do zdaj opravljene večinoma na naglašanih in čistih samoglasnikih, pozornost bo treba preusmeriti tudi na soglasnike in pa predvsem na prosta govornjena besedila, ki niso studijsko posneta. Naravni govor bo treba raziskati podrobneje.

Govorna zbirka, ki so jo ustvarili na Amebisu, d. o. o., je lahko zelo pomembna za vnaprejšnje raziskave. Je verjetno ena največjih govornih zbirk, transkribirana na ožji akustični ravni (raven po Zemljarič Miklavčič). Ko bomo imeli dokončane še samodejno označene posnetke, se bodo lahko dolžine določenih glasov preprosto statistično primerjale, temeljile bodo na podlagi več tisoč primerov (žal le dveh govorcev).

V prihodnosti bo predvsem pomembno razumevanje sistema jezika z vidika računalništva in informatike, kjer so potrebna drugačna in bolj sistematična pravila. Ta nam bodo omogočala pravilno digitalizacijo jezika, ki bo slovenščino še bolj približala uporabniku, nove izboljšane jezikoslovne tehnologije pa bodo jezikoslovcem omogočile hitrejše, natančnejše analize in lažje ter boljše opravljanje dela (npr. lekture).

10 Viri in literatura

1. Marija Golden. *O jeziku in jezikoslovju*. Ljubljana: Filozofska fakulteta, Oddelek za primerjalno in splošno jezikoslovje, 2008.
2. Peter Holozan, Uporaba glagolskih predlog pri strojnem prevajanju. *Jezikovne tehnologije: zbornik B 7. mednarodne multi-konference Informacijska družba IS 2004, od 9. do 15. oktober 2004*. Ljubljana: Institut „Jožef Stefan“, 2004. 128–129.
3. B. H. Juang. L. R. Rabiner. *Automatic speech recognition – A brief history of the technology development*. Elsevier Encyclopedia of Language and Linguistics, Second Edition, 2005. http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf. (Dostop 20. 5. 2013.)
4. Matjaž Rodman, Problemi pri vključevanju ekspertnega lingvističnega znanja v akustično modeliranje. *Jezikovne tehnologije: zbornik B 7. mednarodne multi-konference Informacijska družba IS 2004, od 9. do 15. oktober 2004*. Ljubljana: Institut „Jožef Stefan“, 2004. 119–123.
5. Simon Rozman, *Lepljenje trajajočih glasov v sintezi govora, dipl.* Ljubljana, 2000.
6. Simon Rozman, *Sinteza govornega signala na osnovi metode HNM, mag.* Ljubljana, 2005.
7. Hotimir Tivadar, Peter Jurgec, Podoba govorjenega slovenskega knjižnega jezika v Slovenskem pravopisu 2001. *Slavistična revija* LI/2 (2003). 203–220.
8. Hotimir Tivadar, Normativni vidik slovenščine v 3. tisočletju – knjižna slovenščina med realnostjo in idealnostjo. *Slavistična revija* LVIII/1 (2010). 105–116.
9. Hotimir Tivadar, Gradivna utemeljenost opisa slovenskega govorjenega jezika. *Izzivi sodobnega jezikoslovja*. Ljubljana: Znanstvena založba Filozofske fakultete, 2010. 53–62.
10. Hotimir Tivadar, Vzpostavitev razmerij med govorom in branjem, recitacijo in igranjem. *Meddisciplinarnost v slovenistiki*. Ljubljana: Znanstvena založba Filozofske fakultete, 2011. 494–495.
11. Hotimir Tivadar, Nevarna razmerja med pisnim in govorjenim jezikom. *Pravopisna stikanja: razprave o pravopisnih vprašanjih*. Ljubljana: Založba ZRC, 2012. 203–211.
12. Hotimir Tivadar, Nove usmeritve pri raziskavah govora s pogledom v preteklost. *Slavistična revija* LX/4 (2012). 587–601.

13. Jože Toporišič, *Slovenska slovnica*. Maribor: Obzorja, 2004.
14. Boštjan Vesnicher, France Mihelič, Nikola Pavešič, Vrednotenje na prikritih Markovovih modelih temelječega sistema za umetno tvorjenje slovenskega govora. *Jezikovne tehnologije: zbornik B 7. mednarodne multi-konference Informacijska družba IS 2004, od 9. do 15. oktober 2004*. Ljubljana: Institut „Jožef Stefan“, 2004. 98–102.
15. Melita Zemljak, Zdravko Kačič, Simon Dobrišek idr., Računalniški simbolni fonetični zapis slovenskega govora, *Slavistična revija* L/2 (2002). 159–169.
16. Jana Zemljarič Miklavčič, *Govorni korpusi*. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta, 2008.
17. Jerneja Žganec Gros, *Samodejno tvorjenje govora iz besedil*. Ljubljana: ZRC SAZU, 2000.
18. Jerneja Žganec Gros, Sinteza slovenskega govora – jezikoslovni vidiki. *37. seminar jezika, literature in kulture: zbornik predavanj*. Ljubljana: Filozofska fakulteta, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik, 2001. 21–31.
19. Jerneja Žganec Gros in France Mihelič in Simon Dobrišek, Govorne tehnologije: Pridobivanje in pregled govornih zbirk za slovenski jezik. *Jezik in slovstvo* IIL/3–4 (2003). 47–59.
20. Andrej Žgank idr., Slovenska govorna in tekstovna baza parlamentarnih razprav za avtomatsko razpoznavanje govora. *Jezikovne tehnologije: zbornik 9. mednarodne multi-konference Informacijska družba IS 2006, 9. do 10. oktober 2006*. Ljubljana: Institut „Jožef Stefan“, 2006. 115–118.

11 Priloge

11.1 Anketa o kakovosti in uporabnosti Govorca 2

Spoštovani,

sem Nejc Robida, študent Filozofske fakultete Univerze v Ljubljani. V diplomski nalogi raziskujem sintezo govora, pri raziskovanju pa bi mi bila v veliko pomoč tudi vaša ocena govora Govorca 2.

Pred vami je anketa o kakovosti sinteze Govorca 2. Poleg ankete pošiljam tudi posnetke dveh različnih besedil z dvema možnima izbirama glasov (Matej in Renato). Vsako besedilo je sintetizirano v treh hitrostih (počasneje, normalno in hitreje). Prvo besedilo je publicistično, drugo pa je skupek različnih situacij in intonacij (zasebni pogovor, sleng, narečje in vprašalna poved). Najprej s križcem v razpredelnici z 1 do 5 ocenite kakovost nekaterih lastnosti sinteze in nato še, prosim, odgovorite na vprašanja.

OCENA	1	2	3	4	5
fonemska ustreznost					
naglasna ustreznost					
barva glasu					
intonacija					
register					
poudarki					
dolžina in mesto premorov					
uporabnost uravnavanja hitrosti govora					

Kaj je bilo napačno izgovorjeno (naglas, intonacija, barva, register, premori)?

Kaj vas je pri Govorcu 2 najbolj motilo?

Bi Govorca 2 uporabljali kot bralnik ali kot del aplikacij tehničnih pripomočkov (dlančniki, mobilni telefoni, zvočne knjige)?

11.1.1 Besedilo 1

Vlada Republike Slovenije je na 15. redni seji seznanila državni zbor s paketom protikriznih ukrepov, ki jih bo sprejela še letos. Razprava je trajala le 2 uri. Vse poslanske skupine so se do načrta opredelile in podale tudi svoje predloge. Med možnimi elementi racionalizacije delovanja države so se tako znašli še nujna reforma zdravstva, pretirana in neuravnovešena regionalizacija in plačni sistem v javnem sektorju. Omenjala se je tudi vnovična reforma dela, saj naj bi bilo pri zdajšnji premalo popuščanja tako s strani delodajalcev kot delojemalcev. Vsi so se strinjali, da je zdaj pravi čas za korenite in pogumne spremembe.

11.1.2 Besedilo 2

Ah, nič ni, samo spi se mi.

A je to tvoj pullover rdeč?

Kva pa teb dons tok dogaja, čist si fuknen.

Kiseu krompir, to je pa krampirjeva župa, nter se pa da mal jisha.

11.2 Zgoščanka s posnetki sintetiziranega govora

IZJAVA O AVTORSTVU

Izjavljam, da je diplomsko delo *Sinteza govora in Govorec 3* v celoti moje avtorsko delo in da so uporabljeni viri in literatura navedeni v skladu z mednarodnimi standardi in veljavno zakonodajo.

Podgora, 2. september 2013

Nejc Robida