

# PROBLEMI SAMODEJNE MERITVE SAMOGLASNIKOV

**Nejc Robida**

I. gimnazija v Celju, Celje

**Hotimir Tivadar**

Filozofska fakulteta, Ljubljana

UDK 811.163.6'342.41:81'271.1

Meritve samoglasnikov v slovenščini so izvajali že mnogi raziskovalci, a so bile vedno opravljene ročno in na manjšem vzorcu glasov. Ko smo v podjetju Amebis, d. o. o., razvijali sintetizator govora, smo veliko časa posvetili tudi govorni zbirki, ki so jo v podjetju posneli s pomočjo RTV Slovenija in njihovih napovedovalcev (spikerjev). Korpus smo zopet uporabili ter na nekaterih samoglasnikih opravili nove strojne meritve formantov in še nekaterih drugih lastnosti fonemov. Ta članek se osredotoča na samo metodologijo meritev, ki smo jih opravili, in na primerjavo rezultatov z rezultati ročno delanih analiz.

samoglasniki, govorna zbirka, formantna analiza, govorjeni knjižni jezik, samodejne fonetične meritve

Slovene vowel measurements have been conducted by many researchers but always manually and on a smaller sample of phonemes. While developing the speech synthesizer in the company Amebis, d. o. o. we devoted a lot of time to the speech corpus recorded by the company with the help of Radio-Television Slovenia and their presenters. The corpus was reused and new automatic machine measurements of formants and other characteristics were made of some vowels. The article focuses on the measurement methodology and the comparison of its results with those derived from manual analyses.

vowels, speech corpus, formant analysis, standard spoken Slovene, automatic phonetic measurements

## 1 Uvod

Slovenska govorjena beseda se v sodobnem slovenističnem raziskovalnem prostoru sooča z mnogo izzivi. Ob najavah novih slovarjev se ponovno vračamo k problematiki norme in kodifikacije knjižnega jezika,<sup>1</sup> predvsem njegove glasovne podobe. Od Ruplove izdaje *Slovenskega pravorečja* bo kmalu minilo 70 let, tako je skrajni čas, da se zopet preveri in ponovno oceni norma govorjene podobe knjižne slovenščine. Za to pa potrebujemo čim več raziskav na podlagi čim obsežnejšega

<sup>1</sup> Zavedamo se polemik o poimenovanju knjižna oz. standardna slovenščina. V tem primeru je bil analiziran govor znanih napovedovalcev, spikerjev Radia Slovenija, ki so opravili zahtevno govorno šolanje na nacionalni RTV (več o normi in (ne)knjižnem v Tivadar 2010). Glede na radijske govorce, bralce, napovedovalce in radijsko tradicijo bi bila smiselna mogoče še uporaba zborna izreka (prim. Šeruga-Prek, Antončič 2007). Odločili smo se za govorjeni knjižni jezik, ker je šlo za branje in dejansko govorjeno podobo knjižnega jezika.

gradiva ter natančno določeno, standardizirano metodologijo zajema, obsega in načina analize zbranih podatkov, ki morajo biti tudi sledljivi. Predvsem temu se bomo posvetili v naslednjih poglavjih.

## 2 Analizirano gradivo

Za samodejno analizo lastnosti samoglasnikov smo uporabljali gradivo, ki ga je podjetje Amebis, d. o. o., pridobilo za delovanje Govorca 3, tj. sintetizator govora.<sup>2</sup> Gradivo je podjetje izbralo tako, da je izbor kombinacij glasov in stavčnih intonacij čim širši. Gre za približno dvajset ur posnetkov, ki so razdeljeni na posamezne odseke. Poleg avdiosignalna (16 bitov, 48 kHz) so posneli še signal utripanja glasilk. Bralc<sup>3</sup> izbranega besedila sta bila dva, moški in ženska, oba profesionalna radijska govorca, tj. napovedovalca oz. spikerja, ki sta bila šolana na Radiu Slovenija. Bralca sta bila primerna za analizo govorjenega knjižnega jezika, ker sta bila uradno šolana in ker smo predvidevali, da z vsakodnevno pojavnostjo v medijskem prostoru sooblikujeta podobo standardne govorjene slovenščine. Ker so besedila brana, so tako bližje že mnogo bolj standardizirani pisni podobi jezika; s tem smo se izognili vplivom nezbornih socialnih zvrsti jezika na knjižni govor in izgovor samoglasnikov.<sup>4</sup> Ker govorjeni jezik »v raziskanosti zaostaja za pisnim jezikom, po eni strani zaradi tradicionalno prestižne vloge knjižnega jezika,<sup>5</sup> po drugi strani pa zaradi akustične pojavnosti, ki otežuje zbiranje gradiva in njegovo analizo« (Zemljarič Miklavčič idr. 2009: 423), se je branje zapisanega besedila<sup>6</sup> v primeru akustične analize samoglasnikov izkazalo kot prednost. Pri tem pa je treba poudariti, da gre v tem primeru tudi za branje izkušenih govorcev, ki bodo brali pripravljena besedila kar najbolj naravno (prim. Tivadar 2004), nendaravnost se namreč pogosto zgodi pri manj veščih, neprofesionalnih bralcih.

## 3 Transkripcija<sup>7</sup>

Celotno posneto gradivo smo transkribirali. Najprej se je transkripcija tvorila s pomočjo samodejne grafemsko-fonemske pretvorbe, ki so jo sestavili na Amebisu, d. o. o., po priporočilih Instituta »Jožef Stefan«, nato se je ročno, z večkratnim pregledom celotnega korpusa, popravljala. Tako smo se izognili človeškim napakam zaradi obsežnosti zapisov in napakam samodejne pretvorbe, ko jezik dokazuje svojo

<sup>2</sup> Več o Govorcu 3 v Robida (2013, 2014).

<sup>3</sup> Govorimo o bralcih, saj je bilo posneto besedilo v govorni zbirki brano.

<sup>4</sup> Razvejanost slovenščine, njenih narečij in pokrajinskih govorov vpliva predvsem na izgovor samoglasnikov pri knjižnem govoru neveščih govorcev, npr. reduciranje, celo izpadni glasov *i*, *a*, *u* (Tivadar 2008: 40).

<sup>5</sup> Tu je treba poudariti, da je verjetno mišljen predvsem knjižni jezik kot pisni.

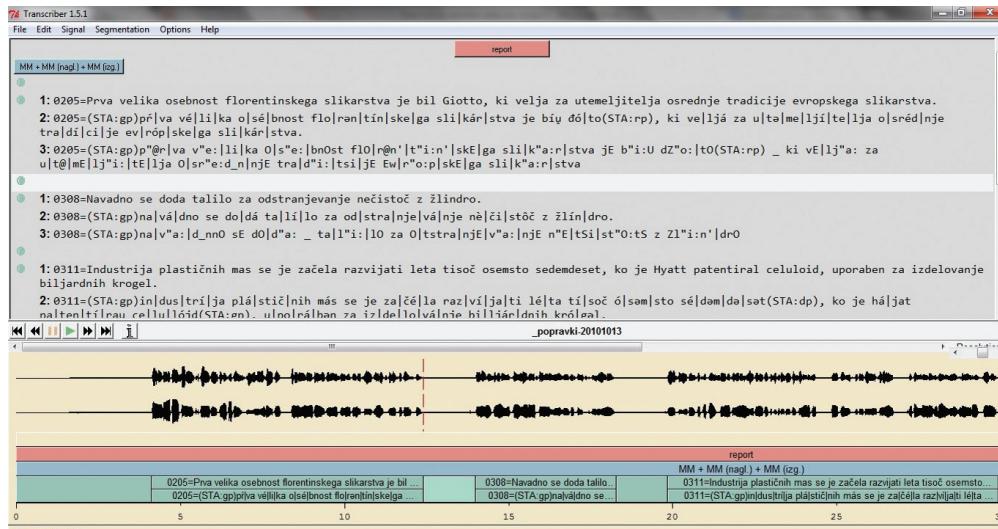
<sup>6</sup> Brano govorno gradivo je sicer v nasprotju z izhodišči izgradnje slovenskega govornega korpusa, ki ima »tri osnovne cilje: zajeti vzorčne primere različnih govorjenih diskurzov v različnih situacijah, zajeti govorjeni diskurz demografsko reprezentativnega vzorca govorcev, zajeti predvsem tiste govorne situacije, v katerih so uporabniki jezika najpogosteje produktivno-receptivno udeleženi« (Zemljarič Miklavčič idr. 2009: 425).

<sup>7</sup> Opis celotnega postopka transkribiranja Amebisove gorovne zbirke je na voljo v Robida 2013.

naravno živost in se pač ne drži vseh pravil.<sup>8</sup> Nekatera pravila pretvorbe smo tudi sproti popravljali. Odseke, ki niso bili jasno prebrani ali so bili v ozadju slišani razni nenačrtovani šumi, smo odstranili. Nekaj glasov smo zaradi poznejše lažje samodejne analize gradiva tudi ukinili, kar je prikazano v Tabeli 1.

Tabela 1: Primeri nekaterih ukinjenih glasov (Robida 2014: 121)

MRPA	IPA	Opis glasu	Primer
b_n	b	favkalni b	območje
b_f	bf	zobnoustnični b	ob fakulteti
d_l	dl	obstranski d	dlan
d_f	ð	favkalni d	odfrleti
w	w	ustničnoustnični šumni zveneči v	vzeti
W	W	ustničnoustnični šumni nezveneči v	vsaj



Slika 1: Delovno okolje popravljanja transkripcij v programu Transcriber

Poleg tega so bili dodani tudi novi simboli, ki jih prej pri podjetju Amebis, d. o. o., niso poznali, na primer grafema za ozka nenaglašena *o* in *e*, saj glede na Toporišičev opis<sup>9</sup> obstajajo samo široki nenaglašeni e-jevski in o-jevski glasovi: »Ozek izgovor glasu *o* smo uvedli pri prihodnjiku bo, čeprav bi moral biti [glede na opis in normo v

8 »V primerih, v katerih se pričakuje izgovor [ij], je velikokrat prišlo do redukcije glasu *i* ali pa se *i* sicer sliši, a se [je pri akustični analizi pokazalo], da nima osnovnega tona in je ustvarjen le s turbulencami v ustni votlini brez pomoči glasilik« (Robida 2013: 33).

9 Toporišič v *Slovenski slovnici* (2004: 56) namreč pravi, da sta »[k]ratka naglašena è in ô ter nenaglašena e in o [...], kot rečeno, zmeraj le široka«.

*Slovenski slovnici* (1976) in *Slovenskem pravopisu* (2001)] izgovorjen široki » (Rybida 2014: 121).<sup>10</sup>

#### 4 Samodejni razrez posnetkov in analiza gradiva

Celotna 20-urna govorna zbirka bralcev se je po končanih pregledih transkripcij poslala na Univerzo v Mariboru, na Fakulteto za elektrotehniko, računalništvo in informatiko, tam so opravili samodejno časovno segmentacijo fonemov. Osnova samodejnega razreza temelji na razpoznavalniku tekočega govora za neodvisnega govorca (angl. ASR – *automatic speech recognition*).<sup>11</sup> Sistem za slovenski jezik je bil naučen na bazi BNSI Broadcast News;<sup>12</sup> gre za 36-urni transkribirani korpus večernih novic RTV Slovenija.

Slovenski jezik sodi zaradi svojih lastnosti (pregibna narava, dvojina, relativno prost vrstni red besed v stavku ...) med težavnejše jezike za razpoznavanje govora [...] Sistem za razpoznavanje tekočega govora UMB Broadcast News je zasnovan na statističnem modeliranju govora. Blokovna shema osnovnih gradnikov sistema je predstavljena na [Sliki 2]» (Andrej Žgank idr. 2006).

Z akustično segmentacijo vhodnega besedila in učenje akustičnih modelov so uporabili kombinacijo Gaussovih modelov (GMM) in prikritih modelov Markova.<sup>13</sup> Akustične modele so zasnovali na grafemski osnovni enoti. Takšen postopek ASR za neodvisnega govorca se je najprej akustično adaptiral še posebej na naša bralca. Nato se je ASR zagnal še v načinu poravnave. Tako je na osnovi transkripcije in fonetičnega slovarja poznal zaporedje fonemov (in s tem akustičnih modelov), najverjetnejše časovne meje pa je ocenil glede na vhodni vektor značilk.<sup>14</sup>

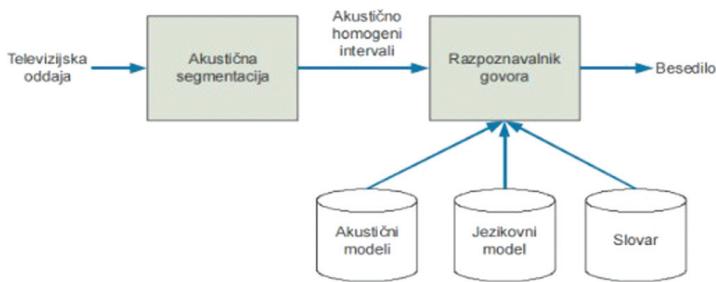
<sup>10</sup> Bralca sta v večini primerov v govorni zbirki, namenjeni Govorcu 3, izgovorila prihodnjik *bo* z ozkim o.

<sup>11</sup> Delovanje razpoznavalnika je povzeto po člankih Žgank idr. (2006) in korespondenci po e-pošti z dne 12. 5. 2014.

<sup>12</sup> Več v članku Žgank idr. 2005.

<sup>13</sup> To so statistični modeli, ki skušajo v našem primeru s strojno vzporedno primerjavo fonetičnega zapisa in glasovnega posnetka ugotoviti uporabne in ponovljive vzorce ter se z njihovo pomočjo naučiti razpoznavanja govora oziroma izgovora glasov, besed ali kar celih povedi, saj se lahko, če imamo to označeno v transkripcijah, naučijo tudi stavčne intonacije.

<sup>14</sup> Za določanje posameznih fonemov na posnetku se je govorni signal analiziral vsakih 10 ms, eno posamezno okno pa je bilo dolžine 25 ms. Na vsakem od teh oken so se opravile meritve, in sicer izračun mel-kestralnih koeficientov in energije signala ter prvi in drugi odvodi teh vrednosti, ki jih skupno imenujemo vektor značilk. Z njihovo pomočjo je lahko nato samodejna analiza poiskala meje med različnimi fonemi in jih določila na 10 ms natančno.



Slika 2: Osnovna zgradba sistema UMB Broadcast News za razpoznavanje govora (Žgank idr. 2006)

## 5 Samodejne meritve samoglasnikov

Po opravljenem razrezu na foneme smo naredili več poskusov ter testirali možnosti za samodejno segmentno in nadsegmentno analizo vseh glasov v našem korpusu.<sup>15</sup> Samodejno smo za vsak glas izračunavali:

- osnovno frekvenco ( $F_0$ ),
- resonance oz. formante ( $F_1, F_2, F_3, F_4$ ),
- rast oz. padanje  $F_0$  (delta  $F_0$ ) – torej potek osnovnega tona oz. intonacijo,
- energijo (E).

Poleg teh podatkov je v analizni dokumentaciji o vseh glasovih v zbirki na voljo še podatek o fonemih, ki stojita pred preučenim glasom in za njim, naklonu stavka, v katerem je preučeni glas, in identifikacijski oznaki povedi. Vse te podatke smo po analizi izvozili v obliko, prilagojeno programu Excel (Slika 3).

<sup>15</sup> Vse dodatne računalniške analize in teste govorne zbirke so naredili v podjetju Amebis, d. o. o. Sodelovali smo z mag. Simonom Rozmanom, ki nam je bil (in je še vedno) v veliko pomoč pri preučevanju govorne zbirke. Postopke teh analiz povzemamo po naši stalni korespondenci z njim v zadnjih letih.

	A	B	C	D	E	F	G	H	I	J	K	L
1	t [ms]	f [Hz]	df [Hz/s]	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	E [dB]	prej. fon.	nasl. fon.	st. naklon	ID povedi
2	130	75,21892	-6,81849	701,941	1260,831	2315,697	3658,479	-21	r	t	gp	0002_00
3	110	68,5716	11,67886	645,8341	1227,214	2306,011	3572,642	-23	v	r	gp	0002_00
4	140	116,5416	327,3471	751,4458	1329,431	2545,439	3702,576	-17	d	r	gp	0003_00
5	130	69,64029	-18,2196	714,7466	1181,39	2333,012	3628,477	-23	v	n'	gp	0003_00
6	60	82,03406	68,7667	644,6242	1219,423	2439,705	3570,87	-19	l	U	gp	0003_00
7	150	79,0066	-94,9057	651,4731	1179,372	2318,452	3587,763	-19	s	d	gp	0003_00
8	130	73,58042	-168,002	675,662	1268,289	2288,24	3617,715	-21	s	l	gp	0004_00
9	120	77,69532	-183,808	670,9545	1229,941	2285,839	3562,322	-20	tS	l	gp	0005_00
10	180	78,82445	-64,8389	764,6013	1260,138	2415,301	3678,572	-21	t	l	gp	0005_00
11	110	97,1496	138,8767	789,4358	1298,113	2200,127	3664,846	-21	j	g	rp	0005_00
12	130	71,29491	-34,0078	777,7198	1182,684	2254,036	3605,957	-23	v	z	rp	0005_00
13	130	94,71311	-434,003	733,1757	1293,905	2209,926	3757,898	-21	j	-	s-	0006_00
14	110	92,32175	-539,933	737,2311	1334,693	2278,105	3698,747	-22	j	-	s-	0006_00
15	110	70,32947	-227,017	741,1215	1151,009	2289,804	3641,405	-23	b	N	gp	0007_00
16	70	81,09869	-163,734	590,1469	1394,592	2444,463	3705,625	-19	l	m	gp	0008_00
17	160	81,67802	46,65213	798,6535	1183,646	2345,81	3624,011	-19	m	m	gp	0008_00
18	80	85,99475	117,1393	756,5751	1206,582	2321,627	3642,26	-20	_	U	gp	0008_00
19	80	58,84899	-101,826	659,4791	1247,882	2233,076	3707,02	-32	v	s	gp	0009_00
20	150	83,10664	-172,999	677,6944	1272,723	2354,03	3702,779	-18	t	l	gp	0011_00
21	130	70,43344	-168,324	721,7737	1323,608	2259,703	3670,845	-23	Z	n	rp	0012_00
22	30	85,49507	-249,821	613,0357	1409,927	2464,836	3559,966	-17	j	U	gp	0013_00
23	150	96,76451	-61,8843	828,5942	1248,192	2265,298	3773,376	-17	m	r	gp	0014_00
24	210	80,385	-87,9936	817,1527	1275,089	2299,76	3759,58	-22	f	l	gp	0014_00
25	140	101,3027	-18,0827	1292,254	2223,552	3764,573	4315,495	-19	k	n	gp	0015_00
26	160	74,49952	-67,4165	816,6502	1183,562	2344,511	3595,512	-24	b	l	gp	0015_00

Slika 3: Delovno okno v programu Excel z analiznimi podatki o posameznih fonemih<sup>16</sup>

## 5.1 Natančnejsi opis postopka meritev

Za izračun osnovne frekvence fonemov smo uporabili posnetek signala, ki prikazuje utripanje glasilk, nato smo za določevanje resonanc oz. formantov izrezali celoten posamičen fonem in ga utežili s Hammingovim oknom.<sup>17</sup> Težišče za računanje je bilo na sredini fonema, s tem je bilo merjenje opravljeno na delu fonema z najverjetnejšim stabilnim stanjem formantov, izognili smo se tudi prehodom v drug glas. Vzorcevalne frekvence smo prilagodili posebej bralcu (10.000 Hz) in bralki (11.000 Hz), kar nam je omogočilo analizo govornega signala do 5.000 Hz za moški glas in 5.500 Hz za ženski glas.<sup>18</sup> Signal smo nato spustili skozi *pre-emphasis filter* (filter za povečevanje magnitud višjih frekvenc<sup>19</sup>). Na koncu smo s pomočjo metode LPC<sup>20</sup> (angl. *linear predictive coding*) pridobili formante govornega trakta.

<sup>16</sup> Na Sliki 3 so vidni rezultati samodejnih meritev dolgega in naglašenega *a*.

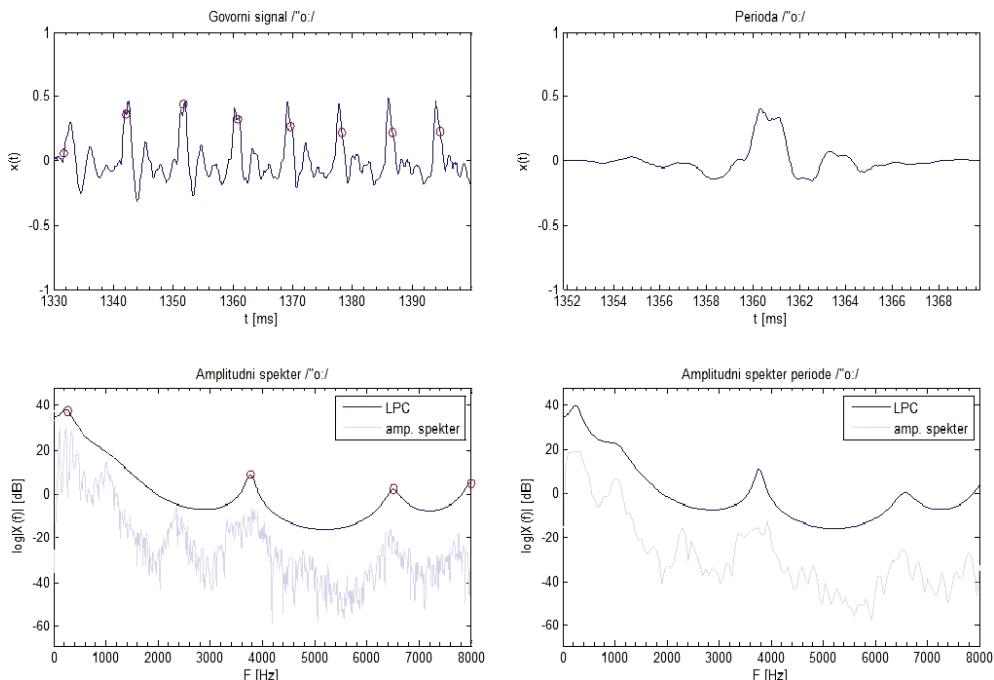
<sup>17</sup> S tem smo v izračunih dali večjo težo sredini fonema in zadušili robove, ki so že lahko prehodi v nov fonem.

<sup>18</sup> Takšne nastavitev smo uporabljali tudi pri ročnih meritvah v programu Praat.

<sup>19</sup> S tem filterom smo poravnali frekvenčni filter, saj energija človeškega glasu na vsako oktavo pada. Tudi to nastavitev smo poenotili z nastavitevami v Praatu, da bi tako dobili čim enotnejši analizi (samodejno in ročno merjenje). Nastavitev filtra je tako 50 Hz +3 dB, 100 Hz +6 dB, 200 Hz +12 dB itn.

<sup>20</sup> LPC je metoda, ki aproksimira spektralno ovojnico s filtrom FIR, za izračun koeficientov filtra uporablja avtokorelačijsko metodo avtoregresivnega modeliranja. Ti koeficienti filtra FIR predstavljajo polinom. Ničle polinoma nastopajo v kompleksno-konjugiranih parih, tako lahko opazujemo samo polovico ničel (ničle z imaginarno komponento večjo ali enako 0). Köt posamezne ničle v polarnem zapisu kompleksnega števila je predstavljal frekvenco kandidata za formant. Frekvence izven intervala [50 Hz, 9950 Hz]

Po naknadnih analizah rezultatov smo ugotovili, da bi morali narediti še načančnejšo analizo samoglasnikov. Odločili smo se, da bomo velikostni red LPC prilagajali posameznemu fonemu vsakega posameznega govorca, saj ima vsak fonem karakteristično postavitev govoril in zato tudi karakteristično število in lokacijo formantov. To smo storili z analizo grafov in spektrograma za vsak posamezni fonem<sup>21</sup> bralca in bralke (Slika 4). Število formantov smo določevali s pomočjo spodnjih dveh spektrogramov, vrisani LPC (modra črta z rdečimi krožci) je bil orientacijska točka, saj je včasih program določil napačno število vrhov na napačnih mestih. Pri dolgem in naglašenem ozkem *o* smo na primer določili štiri vrhove, ki so pomembnejši za naše meritve. Prvi je pri okoli 400 Hz, drugi 1.000 Hz, tretji 2.500 Hz in četrti pri okoli 3.700 Hz. Ti vrhovi predstavljajo tudi približno določene formante  $F_1$ ,  $F_2$ ,  $F_3$  in  $F_4$ . Za dolgi in naglašeni ozki *o* smo zato določili LPC-10 ( $4 \times 2 + 2$ ).<sup>22</sup>



Slika 4: Graf in spektrogram dolgega in naglašenega ozkega *o* (bralka)

## 5.2 Primer meritve dolgega in naglašenega *i*

Zelo dobre poskusne meritve, ki so precej skladne z ročnimi meritvami, smo dobili pri samodejnih meritvah formantov dolgega in naglašenega *i*.<sup>23</sup> Za dolgi naglašeni *i*

pri bralcu in [50 Hz, 10950 Hz] pri bralki ali s širšo pasovno širino od 400 Hz smo izločili, frekvence, ki so ostale, pa so predstavljale formante.

<sup>21</sup> Pri vsakem bralcu smo preučili 59 fonemov. Ločeno smo preverjali glasove, ki so bili (ne)naglašeni, dolgi, kratki, (ne)zveneči idr.

<sup>22</sup> Število 4 predstavlja štiri vrhove oz. formante. Ker je vsak formant v LPC predstavljen z dvema poloma, smo to število pomnožili z 2 in prišeli še dva pola za robne primere.

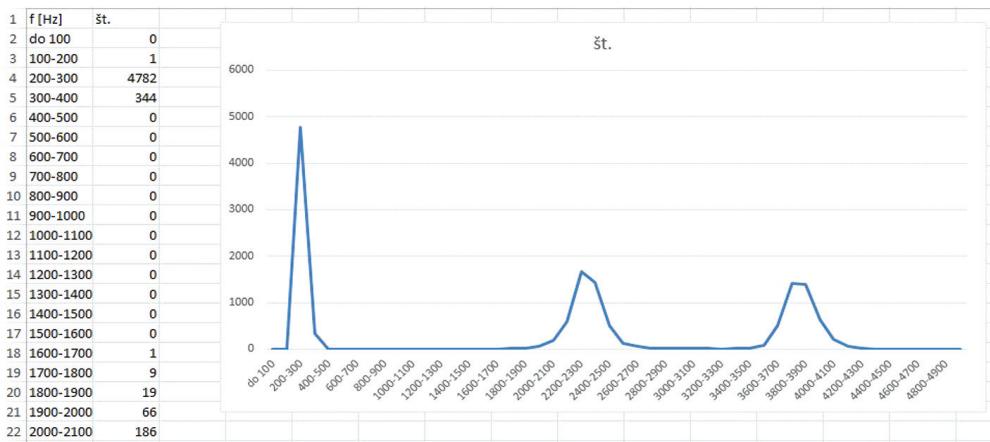
smo imeli pri bralcu na voljo 5.137 primerkov izrezanih fonemov. Ker smo imeli na začetku v programu Excel meritve zaradi omejitev samodejnih analiz formante  $F_1$ ,  $F_2$  in  $F_3$  pomešane (Slika 5), smo to rešili s pomočjo histograma in tabele.

	A	B	C	D	E	F	G	H	I	J	K	L
1	t [ms]	f [Hz]	df [Hz/s]	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	E [dB]	prej. fon.	nasl. fon.	st. naklon	ID povedi
2	60	156,7706	107,5528	229,6131	2442,048	3804,954		-20 t	s	gp	0002_00	
3	100	115,1618	192,2966	242,5554	2309,511	3626,854		-15 t	l	gp	0002_00	
4	100	105,4662	223,1806	250,7752	2244,325	3741,886		-17 l	l	gp	0002_00	
5	100	75,25642	-183,762	283,8516	2379,427			-21 l	n	gp	0002_00	
6	90	125,0483	-274,056	273,6335	2295,8	3774,588		-15 r	"o:	gp	0003_00	
7	30	96,18942	355,7301	265,0429	3714,745			-19 b	U	gp	0004_00	
8	50	134,8179	48,32244	241,7373	2287,837	3784,407		-19 t	s	gp	0004_00	
9	50	111,5452	-85,9477	239,5914	2265,729	3722,713		-16 t	r	gp	0004_00	
10	130	79,15852	-134,642	250,5722	2320,414	3873,138		-22 r	j	s-	0005_00	
11	50	88,55861	-472,354	264,1591	2200,465	3740,395		-16 t	r	s-	0007_00	
12	60	98,62945	137,1213	251,0729	2248,159	3711,307		-16 b	l	gp	0007_00	
13	50	88,55301	-603,361	255,2056	2217,798	3692,083		-18 t	r	gp	0007_00	
14	150	101,1529	198,9844	257,0575	2375,683	3677,136		-28 n	n	gp	0007_00	
15	60	114,1789	541,1888	252,2318	2399,952	3756,787		-14 n	b	gp	0008_00	
16	80	114,8213	140,7646	234,1402	2269,195	3734,041		-17 tS	t	gp	0013_00	
17	90	136,5335	-44,4645	313,4347	2495,241			-18 n	i	pp	0013_00	
18	60	84,74576	5,09E-12	268,9063	2172,834	3666,028		-20 l	l	pp	0013_00	
19	100	163,7155	153,0943	215,1625	2378,49	3722,88		-20 i	t	gp	0014_00	
20	100	121,4336	191,7414	244,0195	2233,304	3707,284		-16 l	f	gp	0014_00	
21	80	88,57458	-17,2845	238,3983	2234,881	3515,742		-16 b	b	gp	0015_00	
22	70	89,62008	-176,89	243,4366	2256,778	3795,365		-19 tS	F	vp	0017_00	
23	50	117,4898	-503,019	248,8697	3734,334			-13 j	G	gg	0018_00	
24	70	117,0577	508,541	234,528	2328,054			-16 b	j	vp	0019_00	
25	150	84,25781	354,5209	244,377	2321,012	3867,769		-24 E	z	vp	0019_00	
26	70	162,1016	372,7361	237,6544	2271,7	3524,106		-18 l	s	gp	0020_00	

Slika 5: Pomešane meritve formantov  $F_1$ ,  $F_2$ ,  $F_3$  in  $F_4$  pri dolgem in naglašenem  $i$  (bralec)

Ustvarili smo tabelo in grafikon, iz katerih smo lahko razbrali, v katerem frekvenčnem razponu se pojavlja največ vrhov (formantov). Na veliki bazi podatkov smo določili, v katerem frekvenčnem območju se pojavljajo  $F_1$ ,  $F_2$ ,  $F_3$  in  $F_4$ . S tem pa smo odstranili še slabe samodejne meritve in slabe primerke izrezanih glasov (Slika 6). Napake so lahko posledica napačne samodejne poravnave, kratke pojavitve fonemov, premočnih vplivov sosednjih fonemov itn.

- 23 V članku smo predstavili meritve le na glasovih  $i$  in  $\text{ə}$ , saj bi bilo drugače delo za ta članek preobsežno. Odločili smo se za srednji in »robni« samoglasnik, zato smo se odločili za  $i$ , ki je visok in sprednji, in  $\text{ə}$ . V prihodnje bo treba preveriti skladnost ročnih in samodejnih meritov še na drugih samoglasnikih.

Slika 6: Histogram dolgega in naglašenega *i* (bralec)

Po določitvi razponov frekvence pri posameznih formantih smo lahko izračunali povprečno frekvenco in standardni odklon izbranega formanta.

	f_min [Hz]	f_max [Hz]	f_povp [Hz]	f_std [Hz]
F1	200	400	266,139659	20,5026
F2	2100	2500	2294,22528	85,58139
F3	3600	4000	3805,77837	88,49305
F4				

Slika 7: Povprečna frekvanca in standardni odklon izbranega formanta pri dolgem in naglašenem *i* (bralec)

Na Sliki 7 lahko razberemo povprečne frekvence posameznih formantov za dolgi naglašeni *i*. Te so primerljive z meritvami drugih raziskovalcev, ki so te meritve opravljali ročno. F<sub>1</sub> ima povprečno frekvenco 266 Hz ( $\pm 8\%$ ), F<sub>2</sub> najdemo pri 2.294 Hz ( $\pm 4\%$ ), F<sub>3</sub> oz. F<sub>4</sub> pri 3.806 Hz ( $\pm 2\%$ ). Pri fonemu *i* je samodejna analiza zaznala samo tri izrazite vrhove. Tako lahko F<sub>3</sub> štejemo tudi kot F<sub>4</sub>, saj se pri ročnih meritvah ta vrh pojavlja kot F<sub>4</sub>. Glede na standardni odklon smo dosegli zelo natančne meritve, a še vedno ostali pri širokem naboru analiziranih primerkov posameznega fonema. F<sub>1</sub> smo določili s povprečjem 5.126 primerkov, za F<sub>2</sub> smo uporabili 4.189 izrezov dolgega naglašenega *i*, za F<sub>3</sub> oz. F<sub>4</sub> pa 3.932 meritve.

Na Sliki 8 so meritve dolgega naglašenega *i*, ki smo jih opravili še pri bralki. Tudi tu smo pri meritvah dobili tri izrazite vrhove.

	A	B	C	D	E
1		f_min [Hz]	f_max [Hz]	f_povp [Hz]	f_std [Hz]
2	F1	200	400	293,029568	30,37278706
3	F2	2500	2900	2679,64947	90,32101344
4	F3	4200	4600	4397,01855	89,71182138

Slika 8: Povprečna frekvanca in standardni odklon izbranega formanta pri dolgem in naglašenem *i* (bralka)

## 6 Primerjava samodejnih in ročnih meritev

### 6.1 Primerjava naših meritev fonema $i$ z meritvami drugih raziskovalcev

Povprečje vseh ročnih meritev, ki so jih do zdaj opravili različni raziskovalci (Tabela 2), pri  $F_1$  je 340 Hz. Tako pridemo v primerjavi z našimi meritvami moškega glasu do 22 % odklona, pri ženskem glasu pa do približno 14 % odklona. Pri meritvah  $F_2$  se ta razlika močno zmanjša. Povprečje vseh ročnih meritev  $F_2$  je 2.171 Hz. Pri moškem glasu novih samodejnih meritev je tako le okoli 6 % odklona, pri ženskem glasu pa 23 % odklona. Tako lahko ugotovimo, da so rezultati naših samodejnih meritev primerljivi z dosedanjimi raziskavami. Odstopanje je sicer vidno, a ni »kritično«.

Tabela 2: Različne meritve samoglasnika  $i^{24}$

Avtor	F1	F2	F3	F4	$F_2 - F_1$	Odklon <sup>25</sup> (%)
Toporišič	382	2116	2797	3552	1734	
Petek-Komar-Šuštaršič	301	2250	2886		1949	
Tivadar (moški glas)	352	2208			1856	
Jurgec (moški glas)	283	2165	2761	3451	1882	±8
Robida-Tivadar (moški glas)	266	2294		3806	2028	±5
Robida-Tivadar (ženski glas)	293	2680		4397	2387	±5

### 6.2 Primerjava ročnih in samodejnih meritev polglasnika

Po primerjavi naših samodejnih meritev z ročnimi meritvami drugih raziskovalcev bomo primerjali še lastne ročne meritve polglasnika in samodejne meritve, ki smo jih dobili od podjetja Amebis, d. o. o. Ta primerjava bo natančnejša, saj bodo meritve narejene na istem gradivu in istih bralcih. Za strojno analizo smo uporabljali skoraj enake nastavitev, kot se uporablajo v programu Praat (opisane že v poglavju 5.1). Edina razlika je v dolžini zajema posameznega analiziranega glasu. Praat uporablja 50 ms okna,<sup>26</sup> mi pa smo zajemali cele foneme, tako smo imeli variabilne dolžine oken. V Tabeli 3 imamo podatke o povprečnih meritvah formantov  $F_1$  in  $F_2$ , med analizirano gradivo smo zajeli naglašene in nenaglašene polglasnike, in sicer z različnih položajev v besedah. Pri bralcu smo analizirali 16 primerkov, pri bralki pa 17. Za vsak formant smo izračunali še povprečja vseh meritev moškega in ženskega glasu.

<sup>24</sup> Prve tri meritve (Toporišič, Petek-Komar-Šuštaršič, Tivadar) smo povzeli po Tivadar 2003, Jurgecove meritve pa so iz Jurgec 2005. Če je bilo možno, smo citirali meritve za moški glas, in sicer meritve dolgega naglašenega  $i$ , drugače smo uporabili povprečje vseh meritev glasu  $i$ .

<sup>25</sup> Odklon je povprečje vseh odklonov pri meritvah formantov ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ ) glasu  $i$ .

<sup>26</sup> Okno določa dolžino in težišče opazovanega odseka govornega signala.

Tabela 3: Ročne meritve polglasnika (povprečje naglašenega in nenaglašenega)

Polglasnik	$F_1$	$F_2$
Bralec (16 analiziranih)	424	1425
Bralka (17 analiziranih)	469	1426
Povprečje	447	1426

Za izračun povprečnih vrednosti formantov  $F_1$  in  $F_2$ , ki smo jih dobili s samodejnimi meritvami, smo ponovili postopek, opisan v poglavju 5.2. Izračunali smo skupna povprečja formantov  $F_1$  in  $F_2$  za naglašeni in nenaglašeni polglasnik pri bralcu in bralki (Tabela 4).

Tabela 4: Povprečja samodejnih meritev polglasnika

Polglasnik	$F_1$	$F_2$
Bralec	393	1336
Bralka	443	1513
Povprečje	418	1425

Primerjava samodejnih meritev z ročnimi (Tabela 5) je pokazala majhna odstopanja, in sicer manj kot 10-odstotna. Rezultat je potrdil pravilnost in zanesljivost strojnih meritev.

Tabela 5: Odstopanja samodejnih meritev od ročnih v odstotkih

Polglasnik	$F_1$ (%)	$F_2$ (%)
Bralec	-8	-7
Bralka	+6	+6
Povprečje	+7	-0,1

Za ročne meritve smo uporabili 33 različnih polglasnikov, pri strojni analizi pa na primer samo pri bralki 6.086 posameznih meritev  $F_1$  in  $F_2$ .

## 7 Zaključek

Glede na vse več možnosti pridobivanja posnetkov in različnih govorcev slovenskega jezika je nujno razmišljati o avtomatizaciji meritev posameznih glasov. V članku smo zato natančno popisali postopek strojnih meritev glasov, te pa smo nato primerjali še z ročnimi. Prvič se je analiza lastnosti samoglasnikov opravila na takoj številčnem gradivu. Rezultati kontrastivne analize ročnega in strojnega merjenja glasov so potrdili primerljivost ter zanesljivost postopkov, ki smo jih uporabljali. Predvsem pa je s tem strojno merjenje na tovrstnih osnovah gotovo lahko dobra podlaga pri nadaljnji fonetično-fonoloških analizah. Seveda nas čaka še veliko dela, s tem člankom pa smo si predvsem potrdili smiselnost nadaljnjih strojnih meritev. Poudariti je treba, da smo imeli pri analizi na voljo samo dva govorca, premalo, da bi lahko govorili o reprezentativnosti, a že posnetih in natančno transkribiranih večurnih

govornih zbirk žal ni veliko na voljo. Za obsežnejšo analizo več govorcev so potrebne nove raziskave in predvsem nova materialna sredstva. Cilj pa bo še naprej ostajal isti – natančen opis jezika, v prvi vrsti nas bo zanimala predvsem normirana slovenščina (govorjeni knjižni jezik), ki bo podlaga tudi nadaljnjam govornotehnološkim raziskavam (sinteza in razpoznavanje govora idr.).

## Viri

- JURGEC, Peter, 2005: Formant frequencies of standard Slovene vowels. *Govor* 22/2. Zagreb: Odjel za fonetiku Hrvatskoga filološkog društva. 127–144.  
<http://www.jurgec.net/publications/formantfreq.pdf>
- ROBIDA, Nejc, 2013: *Sinteza govora in Govorec 3*. Diplomska naloga. Ljubljana: Filozofska fakulteta.
- ROBIDA, Nejc, 2014: Sinteza govora in Govorec 3. Hotimir Tivadar (ur.): *Prihodnost v slovenskem jeziku, literaturi in kulturi. 50. SSJLK*. Ljubljana: Znanstvena založba Filozofske fakultete. 117–124.  
[http://www.centerslo.net/files/file/ssjlk/50\\_SSJLK/Robida.pdf](http://www.centerslo.net/files/file/ssjlk/50_SSJLK/Robida.pdf)
- RUPEL, Mirko, 1946: *Slovensko pravorečje: navodila za zborno ali knjižno izreko*. Ljubljana: Državna založba Slovenije.
- ŠERUGA-PREK, Cvetka, ANTONČIČ, Emica, 2007: *Slovenska zborna izreka: priročnik z vajami za javne govorce*. Maribor: Založba Aristej.
- TIVADAR, Hotimir, 2003: *Govorjena podoba slovenskega knjižnega jezika – pravorečni vidik*. Magistrska naloga. Ljubljana, Praga.
- TIVADAR, Hotimir, 2004: Fonetično-fonološke lastnosti samoglasnikov v sodobnem knjižnem jeziku. *Slavistična revija* 52/1. [31]–48.
- TIVADAR, Hotimir, 2008: *Kakovost in trajanje samoglasnikov v govorjenem knjižnem jeziku*. Doktorska disertacija. Ljubljana, Praga.
- TIVADAR, Hotimir, 2010: Normativni vidik slovenščine v 3. tisočletju – knjižna slovenščina med realnostjo in idealnostjo. *Slavistična revija* 58/1. 105–116.  
<http://www.dlib.si/details/URN:NBN:SI:DOC-BHCJ5UO8>
- TOPORIŠIČ, Jože, 2004: *Slovenska slovnica*. Maribor: Obzorja.
- ZEMLJARIČ MIKLAVČIČ, Jana, 2008: *Govorni korpusi*. Ljubljana: Znanstvena založba Filozofske fakultete.
- ZEMLJARIČ MIKLAVČIČ, Jana, STABEJ, Marko, KREK, Simon, ZWITTER VITEZ, Ana, 2009: Kaj in zakaj v referenčni govorni korpus slovenščine. Marko Stabej (ur.): *Infrastruktura slovenščine in slovenistike. Obdobja* 28. Ljubljana: Znanstvena založba Filozofske fakultete. 423–428.  
[http://www.centerslo.net/files/file/simpozij/simp28/Zemlj\\_Stab\\_Krek\\_Zwitter.pdf](http://www.centerslo.net/files/file/simpozij/simp28/Zemlj_Stab_Krek_Zwitter.pdf)
- ŽGANK, Andrej, VERDONIK, Darinka, ZÖGLING MARKUŠ, Aleksandra, KAČIČ, Zdravko, 2005: BNSI Slovenian Broadcast News database – speech and text corpus. *Interspeech Lisboa 2005: proceedings of the 9<sup>th</sup> European conference on speech communication and technology*. Bonn: Universität, Institut für Kommunikationsforschung und Phonetik.  
[http://www.lrec-conf.org/proceedings/lrec2014/pdf/710\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/710_Paper.pdf)
- ŽGANK, Andrej, ROTOVNIK, Tomaž, SEPESY MAUČEC, Mirjam, KAČIČ, Zdravko, 2006: Osnovna zgradba razpoznavalnika slovenskega tekočega govora UMB Broadcast News. *Jezikovne tehnologije: zbornik 9. mednarodne multi-konference Informacijska družba IS 2006*, 9. do 10. oktober 2006. Ljubljana: Institut »Jožef Stefan«. 99–118. [http://nl.ijs.si/isjt06/proc/19\\_Zgank\\_1of2.pdf](http://nl.ijs.si/isjt06/proc/19_Zgank_1of2.pdf)