SEGMENTAL PROSODIC PARAMETERS IN STANDARD SLOVENE

F. Miheli , J. Gros, T. Srebot-Rejec*, N. Paveši
Artificial Perception Laboratory
Faculty of Electrical Engineering
University of Ljubljana
Tr aška 25, SI-1000 Ljubljana, Slovenia
e-mail: nejka@fe.uni-lj.si
**Department of Comparative and General Linguistics*
Faculty of Arts of the University of Ljubljana
Ašker eva 2, SI-1000 Ljubljana, Slovenia

ABSTRACT

For the Slovene language, only a few studies of segmental prosodic parameters performed on a representative speech corpus have been reported so far. The most comprehensive and carefully designed is the work of T. Srebot-Rejec, providing an insight into Slovene vowel duration and fundamental freqency on the segmental as well as on the suprasegmental level (under the influence of sentence intonation) [1].

We used the study of T. Srebot-Rejec as a reference and repeated some of her measurements of vowel duration and fundamental frequency, with a new speaker. We wanted to check whether we could verify her findings and explanations on the nature of word accent and vowel duration in Standard Slovene. On the other hand, we were looking for a relevant parameter set for prosody prediction in the SQEL Slovene Speech Synthesis System S5 [2,3].

1. INTRODUCTION

A number of studies suggest that prosody has great impact on the intelligibility and naturalness of speech perception. Only the proper choice of prosodic parameters, given by sound duration and intonation contours, enables the production of natural-sounding high quality synthetic speech.

Regardless of whether the speech units are words, syllables or phonetic segments, contextual effects on phone duration and fundamental frequency are complex and involve multiple factors [4,5,6]. Prosody measurements on representative speech corpora are essential for extracting prosodic parameters to build relevant prosody models for a speech synthesis system.

For Standard Slovene, only a few studies of segmental prosodic parameters performed on a representative speech corpus have been reported so far [7,1,8,9]. The most comprehensive and carefully designed is the work of T. Srebot-Rejec, providing an insight into Slovene vowel duration and fundamental freqency on the segmental as well as on the suprasegmental level (under the influence of sentence intonation) [1].

We used the study of T. Srebot-Rejec as a reference and repeated some of her measurements of vowel duration and fundamental frequency (F0), with a new speaker. We wanted to check whether we could verify her findings and explanations on the nature of word accent and vowel duration in Standard Slovene. On the other hand, we were looking for a relevant parameter set for prosody prediction in the SQEL Slovene Speech Synthesis System S5 [2,3].

2. PROSODY PREDICTION IN S5

For prosody prediction in the S5 speech synthesiser we use a two-level approach for both duration and F0 modelling. The prosody prediction process consists of four phases:

- intrinsic duration assignment,
- extrinsic duration assignment,
- modelling of the intra word F0 contour and
- assignment of a global intonation contour.

2.1. Duration Modelling

In our two-level duration model the levels correspond to the two levels of durational control [10]: the extrinsic and the intrinsic one. Units of word length are said to have a set of *intrinsic* durations, stored in our mental lexicon. As these units are integrated into larger entities, such as phrases, they get stretched and squeezed in accordance to larger speech demands, which correspond to an *extrinsic* level of durational control.

We first determine the words' intrinsic duration, taking into account factors, relating to phone segmental duration, such as: segmental identity, phone context, syllabic stress and syllable type: open or closed syllable.

Further, the extrinsic duration of a word is predicted, according to higher-level rhythmic and structural constraints of a phrase, operating on the syllable level and above. Here the following factors are considered: the chosen speaking rate, the number of syllables within a word and the word's position within a phrase, which can be phrase initial, phrase final or nested within a phrase.

Finally, the intrinsic segment duration is modified, so that the entire word acquires its predetermined extrinsic duration. A method for segment duration prediction was developed, which adapts a word with an intrinsic duration to the desired extrinsic duration, taking into account how stretching and squeezing apply to duration of individual segments [3,11].

To apply the two-level approach, different aspects of phone and syllable duration have to be measured, e.g. intrinsic phone duration, syllable duration, the influence of speaking rate on

duration of speech units. This paper concentrates mainly on measurements of intrinsic phone duration in Standard Slovene.

2.2. F0 Modelling

Since Standard Slovene is considered to be a pitch accent language [1], special attention was paid to the prediction of tonemic accents for individual words.

First intrinsic vowel fundamental frequencies were determined according to previous measurements as suggested in [6], creating the F0 backbone. Each stressed word was assigned one of the two tonemic accents, characteristic for the Slovene language. The acute accent is mostly realised by a rise on the posttonic syllable, while with the circumflex the tonal peak usually occurs within the tonic. Five typical F0 patterns were chosen from the variety of F0 patterns described in [1]. Finally a linear interpolation between the defined F0 values was performed.

We used a relatively simple approach for prosody parsing and the automatic prediction of Slovene intonational prosody which makes no use of syntactic or semantic processing, but rather uses punctuation marks and searches for grammatical words, mainly conjunctions which introduce pauses. In the paper we describe measurements of intrinsic vowel fundamental frequency which were later used to initialize the F0 contour for a word.

3. SPEECH CORPUS

3.1. Vowel Duration and F0

A speech database consisting of logatoms, carefully chosen by a phonetician, was recorded in order to study different effects on vowel duration and F0, which operate on the segmental basis in Standard Slovene. The same speaker FM was used who had previously recorded the diphone database and the continuous speech corpus for consonant duration measurements.

To eliminate the influence of adjacent consonants and to measure vowel duration in ideal conditions, the same logatoms - artificial nonsense words were used as in the previous study of T. Srebot-Rejec [1]. The target vowels were studied in logatoms of different length and syllable structure:

'V:CV	long stressed vowel in an open syllable, followed by an unstressed syllable
CV'CV:CV	long stressed vowel in an open syllable, preceded and followed by an unstressed syllable
CV'CV:C	stressed long vowel in a finally stressed closed syllable, preceded by an unstressed syllable
CV'CVC	stressed short vowel in a finally stressed closed syllable, preceded by an unstressed syllable

The target vowels are underlined. As in [1] the logatoms were embedded in mid sentence position so as to minimize the influence of sentence intonation.

3.2. Consonant Duration

A large continuous speech database was recorded to study phone and phoneme group duration in Standard Slovene with the emphasis on consonant duration in various phonetic contexts.

We opted for a relatively long text of 266 sentences derived from the Slovene speech database GOPOLIS [12], covering the domain of air timetable information retrieval.

The male speaker FM was recorded again. The speech material was initially labelled using a Hidden Markov model speech recogniser for the Slovene language in forced segmentation mode. The obtained labels were manually corrected using a special visual interface we developed for viewing, editing and labelling speech signals.

When pronuncing the text, the speaker kept the speaking rate rather constant, as it can be seen from Table 1, showing the phone duration variation. Phone duration variation was evaluated for a given speaking rate by averaging phone duration differences for words, which occured in the corpus several times, in the same phonetic environment and in the same type of phrase.

An average absolute phone duration difference of 5.3 ms with a standard deviation of 8.2 ms was obtained for different realisations of the intial part of the phrase *Ob kateri uri* ..., meaning *At what time* ... for the normal speaking rate (Table 1).

speech rate	average absolute phone duration difference [ms]	standard devation [ms]
normal	5.3	8.2
fast	4.0	6.4
slow	13.8	20.6

Table 1. Phone duration variation for the phrase *Ob kateri uri* given in form of average absolute phone duration difference and the standard deviation.

4. INTRINSIC VOWEL DURATION AND FUNDAMENTAL FREQUENCY IN LOGATOMS

Vowel duration and fundamental frequency were studied in different types of logatom syllables: stressed and unstressed, open and closed, word initial and word final. In most cases, observations given in [1] were confirmed.

4.1. Vowel duration

As in [1], vowel duration in logatoms was studied in neutral intonation position in 2 ways:

- paradigmatically – intrinsic duration
- syntagmatically – in stressed and unstressed syllables

Average vowel duration for stressed vowels in barytones and oxytones is given in Table 2. From Table 3 it follows that unstressed vowels in prestressed position are in average 27% shorter in comparison to unstressed vowels in poststressed syllables. Their duration is limited by the following consonant. Unstressed vowels in prestressed position are in average 51% shorter in comparison to stressed vowels.

4.2. Fundamental frequencies for Slovene vowels

Tables 4 and 5 give F0 measurement results for stressed, prestressed and poststressed vowels.

Vowels in the stressed syllables are mainly higher in than those in poststressed syllables. Speaker FM pronounced all the logatoms with a circumflex accent.

	barytones		oxytones	
	stressed vowel duration [ms]		stressed vowel duration [ms]	
vowel	'(C)V:CV	CV'CV:CV	CV'CV:C	CV'CVC
i :	94	80	93	
i"				62
E:	165	146	148	
E"				82
e:	123	191	90	
a"				111
a:	144	171	227	
O"				96
o:	133	243	218	
O:	124	233	144	
u"				102
u:	108	152	104	
average	127	174	146	90

Table 2. Average duration of long and short stressed vowels in barytones and ocsytones.

	prestressed syllable		poststressed syllable	
	vowel duration [ms]		vowel duration [ms]	
vowel	CV'CV:CV	CV'CVCV CV'CV:CV	'CV:CV	CV'CV:CV
i	36	51	52	70
E	93	86	116	126
a	78	72	113	107
O	111	80	119	112
u	78	68	123	120
average	79	71	105	107

Table 3. Average duration of unstressed vowels in prestressed and poststressed syllables.

	barytones		ocsytones	
	F0 of stressed vowels [Hz]		F0 of stressed vowels [Hz]	
vowel	'(C)V:CV	CV'CV:CV	CV'CV:C	CV'CVC
i :	180	207	205	
i"				186
E:	139	145	143	
E"				187
e:	170	158	182	
a"				157
a:	153	129	130	
O"				126
o:	181	169	152	
O:	137	136	140	
u"				177
u:	192	187	169	
average	165	162	160	169

Table 4. Average F0 in long and short stressed vowels in barytones and ocsytones.

	prestressed syllable		poststressed syllable	
	vowel fundamental frequency [Hz]		vowel fundamental frequency [Hz]	
vowel	CV'CV:CV	CV'CVCV CV'CV:CV	'CV:CV	CV'CV:CV
i	141	140	115	111
E	132	147	105	117
A	102	117	98	89
O	143	128	97	99
U	138	128	106	107
Average	131	132	114	105

Table 5. Average fundamental frequency of unstressed vowels in prestressed and poststressed syllables.

5. PHONE DURATION IN CONTINUOUS SPEECH

On the GOPOLIS continuous speech database average stressed and unstressed vowel durations were measured. Consonant duration was measured in CC and VCV clusters (Table 6).

Consonant duration [ms]	VCV	CC
p	68	44
t	84	64
k	68	65
b	67	60
d	53	50
g	98	55
s	103	54
S	110	58
z	59	57
Z	62	58
dZ	69	66
m	62	54
n	45	34
h	98	33
l	96	50
v	59	48
r	43	45
j	42	37
f	86	85
ts	131	100
tS	84	60

Table 6. Consonant duration in CC clusters and in VCV sequences.

Consonants in CC clusters are in average 23% shorter in comparison to single consonants in VCV sequences.

6. CONCLUSION

Measurements of segmental prosodic parameters in Standard Slovene speech are presented.

The gained prosodic parameters were directly applied in the prosody predictions models for segmental prosodic parameters.

The observations on vowel duration and fundamental frequency mainly conformed to those of T. Srebot-Rejec presented in [1]. More measurement results are presented in [3] where they are discussed in detail.

ACKNOWLEDGMENTS

This work was funded by the Commission of the European Community under COP-94 contract No. 01634 (SQEL) and by the Slovenian Ministry of Science and Technology.

REFERENCES

[1] Srebot-Rejec, T. 1988. *Word Accent and Vowel Duration in Standard Slovene: An Acoustic and Linguistic Investigation*. Slawistische Beiträge. Band 226. Verlag Otto Sagner, München.

[2] Gros, J., Paveši , N. and Miheli , F. 1996. A text-to-speech system for the Slovenian language. *Proceedings of the EUSIPCO'96*. Trieste. 1043-1046.

[3] Gros, J. 1997. Converting Slovenian Text into Speech. *PhD Thesis*. University of Ljubljana. 1997. (in Slovene).

[4] van Santen, J.P.H. 1993. Timing in text-to-speech systems. *Proceedings of the EUROSPEECH'93*. Berlin. 1397-1404.

[5] Klatt, D.H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*. Vol. 59. 1209-1221.

[6] Ohno, S. and Fujisaki, H. 1995. *A method for quantitative analysis of the local speech rate*. Proceedings of the EUROSPEECH'95. Madrid, 421-424.

[7] Topori i , J. 1984 . *Slovenska slovnica*. Maribor. (in Slovene).

[8] Petek, B., Šuštarši , R . and Komar, S. 1996. An acoustic analysis of contemporary vowels of the Standard Slovenian language. *Proceedings of the ICSLP'96*. Philadelphia. 33-136.

[9] Zemljak, M. 1998. Dol ine samoglasnikov v razli nih polo ajih besed in povedi. *Proceedings of the ERK'98*. Portoro , Slovenia. 197-200.

[10] Ferreira, F. 1993. Creation of prosody during sentence production. *Psychological Review*. No. 2. 233-253.

[11] Gros, J., Paveši , N. and Miheli , F. 1997. Speech timing in Slovenian TTS. *Proceedings of the EUROSPEECH'97*. Rhodes.

[12] Dobrišek, S., Gros, J., Miheli , F., Pepelnjak, K. and Ipši , I. 1996. GOPOLIS. *Proceedings of the 2nd SDRV Workshop on Speech and Image Understanding*. Ljubljana. 37-46.