# SYLLABLE AND SEGMENT DURATION AT DIFFERENT SPEAKING RATES IN THE SLOVENIAN LANGUAGE

*J. Gros, N. Pavešić, F. Mihelič*
Artificial Perception Laboratory
Faculty of Electrical Engineering
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia
e-mail: nejka@fe.uni-lj.si

## ABSTRACT

Speech timing at different speaking rates was studied for the Slovenian language and the results were applied in the two level duration prediction model in the Slovenian text-to-speech system *S5* [1].

In order to provide the synthesiser with the possibility to pronounce input text with several speaking rates, tests were made to study the impact of speaking rate on syllable duration and duration of individual phonemes and phoneme groups for the Slovenian language.

## 1. INTRODUCTION

Similarly to Epitropakis [2], we use a two-level approach to duration modelling, which first determines the words' intrinsic duration, taking into account factors, relating to phoneme segmental duration, such as: segmental identity, phoneme context, syllabic stress and syllable type: open or closed syllable [3].

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, taking into account how stretching and squeezing apply for individual phonemes [4]. It is to be noted, that stretching and squeezing does not apply to all segments equally. Stop consonants, for example, are much less subject to temporal modification than other types of segments, such as vowels or fricatives.

To apply the two level approach, different aspects of phoneme and syllable duration have to be measured, e.g. intrinsic phoneme duration. Results of perception experiments show that tempo variation contributes significantly to the perceived naturalness of speech [5]. In order to enable the synthesiser to pronounce input text with several speaking rates, tests were made to study the impact of speaking rate on syllable duration and duration of individual phonemes and phoneme groups.

## 2. SPEECH DATABASE

### *2.1 Vowel Duration*

A speech database consisting of logatoms, carefully chosen by phoneticians [3], was recorded in order to study different effects on vowel duration, which operate on the segmental basis.

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 [3]. The target vowels were studied in logatoms of different length and syllable structure:

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

The target vowels are underlined. The logatoms were embedded in the middle of phrases so as to minimize the influence of sentence intonation [3].

### *2.2 Phoneme and Syllable Duration at Different Speaking Rates*

A large continuous speech database was recorded to study the impact of speaking rate on syllable duration and duration of phonemes and phoneme groups in the Slovenian language.

When reading the same text at different speaking rates, it is possible to obtain phoneme realisations that differ only in duration. Thus context, stress and all other factors are kept identical to every realisation of the sentence. As a result, pair-wise comparisons of phoneme duration can be made.

We opted for a relatively long text of 172 sentences derived from the Slovenian speech database GOPOLIS [6], covering the domain of airflight information services. A male speaker was instructed to pronounce the

same material at different rates: at a normal rate, very slow and as fast as possible. Reading the text took:

| | |
|---|---|
| normal rate | 7 minutes 32 seconds |
| fast rate | 5 minutes 45 seconds |
| slow rate | 12 minutes 55 seconds |

As in [7], the speech material was initially labelled using a Hidden Markov model speech recogniser for the Slovenian language in forced segmentation mode. The obtained labels were manually corrected.

## 3. INTRINSIC PHONEME DURATION

Vowel duration was studied in different types of logatom syllables: stressed and unstressed, open and closed, word initial and word final. Observations on vowel duration given in [3] were confirmed. Additionally, consonant duration was measured in CC and VCV clusters in the normal rate continous speech database.

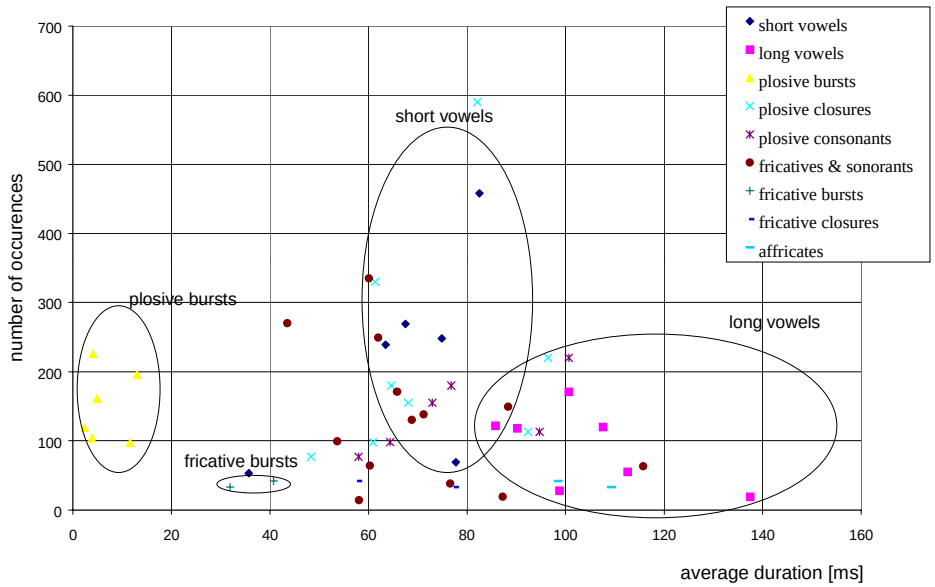


Figure 1: *Duration of phonemes according to their phoneme groups at normal speaking rate.*

Figure 1 shows how phoneme durations for related phonemes belonging to the same group tend to cluster. The clusters are marked by hand. No automatic clustering procedure was applied.

As a result of this study, initial values for inherent phoneme duration when calculating durational parameters were determined.

## 4. PHONEME DURATION AT DIFFERENT SPEAKING RATES

The effect of speaking rate on phoneme duration was studied in a number of ways. An extensive statistical analysis of lengthening and shortening of individual phonemes, phoneme groups (nasals, liquids, plosives, fricatives) and phoneme components (closures, bursts) was made, the first of this kind for the Slovenian language.

Pair-wise comparisons of phoneme duration were calculated. Average mean duration differences and standard devations were computed for pairs of phonemes pronounced at different speaking rates.

Prior to the comparison, phoneme duration was normalised to the corresponding normal rate phoneme duration. Pairs were first composed of normal and slow rate phonemes, and later of fast and normal rate phonemes. Figures 2 and 4 give the results of these pair-wise comparisons and show in what extent the average phoneme duration when speaking or slow or fast increases or reduces with respect to its normal rate duration.

Closures of plosives change but slightly and maintain almost the same duration regardless of the speaking rate. Affricate closures exhibit an interesting behaviour, since they lengthen considerably, whereas they do not shorten at all.

The opposite holds for affricate bursts, together with their corresponding fricatives. In the fricative group, voiced fricatives change more than unvoiced ones. Short vowels, contrary to long vowels, increase more in duration when speaking slower than they shorten when speaking faster.

From these observations we may draw a conclusion: phonemes or phonemic components, which are considered as short by nature (except for bursts of plosives), increase more in duration at a slow rate than they shorten at a fast rate. The opposite holds for affricates and long vowels.

## 5. SYLLABLE DURATION AND ARTICULATION RATE

Articulation rate, expressed as the number of syllables per second [8], excluding silences and filled pauses, was studied for the three different speaking rates for different word positions within a phrase.

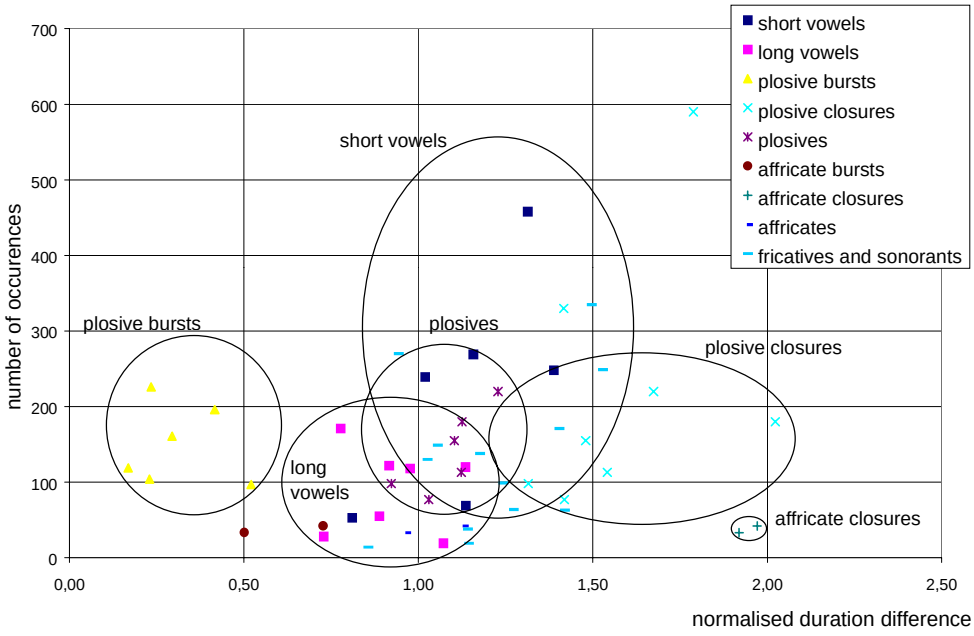


Figure 2: *Pair-wise analysis: normal rate - slow rate. Normalised mean duration difference for pairs of phoneme realisations in the phoneme group context.*

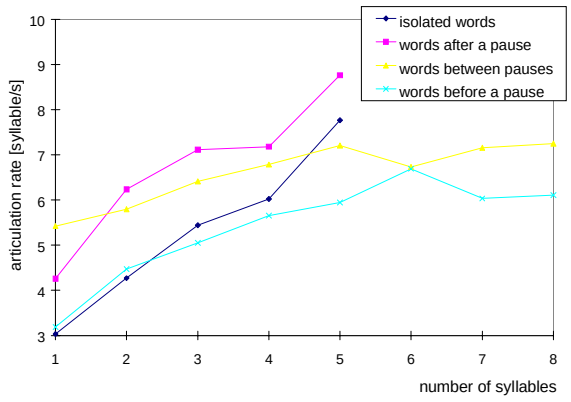


Figure 3: *Articulation rate in number of syllables per second, is given for different word positions.*

Figure 3 shows articulation rate, given in number of syllables per second, plotted as a function of word length, given in number of syllables, and the word position in a phrase. The obtained values apply for normal speaking rate.

The articulation rate immediately after pauses is higher than the one prior to pauses. This prepausal lengthening may be attributed as a slowing down of the speech in anticipation of a pause [8]. The articulation rate increases with longer words as average syllable duration tends to decrease with more syllables in a word.

We observed that in case atona are associated to their neighbouring words, articulation rate adopts a quasi-logarithmic contour, which can be described parametrically, as in [9]. Isolated words and those following a pause differ from this rule for words with more than four syllables, of which only a few realisations were available.

## 6. CONCLUSION

Measurements of different durational parameters of Slovenian phonemes and syllables are presented and discussed. The results were directly applied in the two level approach for duration prediction in the Slovenian speech synthesiser *S5*.

A comprehensive perceptual evaluation of the quality of the synthetic speech was performed, according to ITU-T Recommendation P.85 [10], describing a method for subjective performance assessment of the quality of speech voice output devices.

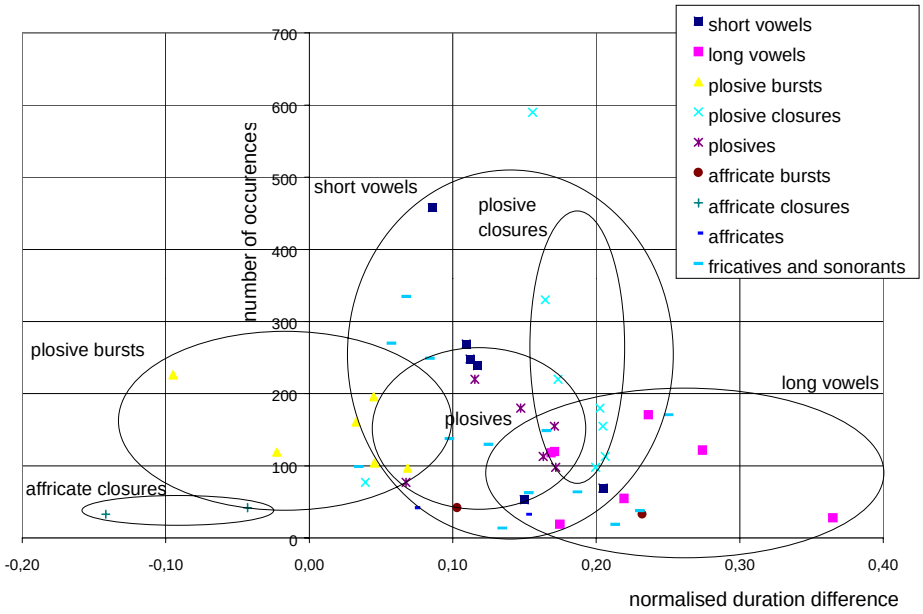


Figure 4: *Pair-wise analysis: normal rate - fast rate. Normalised mean duration difference for pairs of phoneme realisations in the phoneme group context*

## ACKNOWLEDGEMENT

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] J.Gros, N.Pavešić, F.Mihelič, "*A text-to-speech system for the Slovenian language*", Proc. EUSIPCO'96, Trieste, pp. 1043-1046, 1996.

[2] G.Epitropakis, D. Tambakas, N. Fakotakis, G. Kokkinakis, *"Duration modelling for the Greek language*", Proceedings of the EUROSPEECH'93, Berlin, pp. 1995-1998, 1993.

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

[4] J.Gros, N.Pavešić, F.Mihelič, "*Speech timing in Slovenian TTS*", Proc. EUROSPEECH'97, Rhodes, 1997.

[5] S. Ohno, H. Fujisaki, "A method for quantitative analysis of the local speech rate", Proc. EUROSPEECH'95, Madrid, pp. 421-424. 1995.

[6] S. Dobrišek, J. Gros, F. Mihelič, K. Pepelnjak, I. Ipšić, "*GOPOLIS: Slovenian speech database of spoken flight information queries*", Proc. 2nd SDRV Workshop on Speech and Image Understanding, Ljubljana, pp. 37-46, 1996,

[7] J.Gros, I.Ipšić, N.Pavešić, F.Mihelič, S. Dobrišek, "*Automatic segmentation of Slovenian diphone inventories*". Proc. COLING'96, Copenhagen, pp. 298-303, 1996.

[8] D.O'Shaughnessy, "*Timing patterns in fluent and disfluent spontaneous speech*", Proc. ICASSP'95, Detroit, pp. 600-603, 1995.

[9] J. Bakran, "A model of the temporal organisation of the Standard Croatian language", PhD Thesis, University of Zagreb, 1994, in Croatian.

[10] ITU, "*A method for subjective performance assessment of the quality of speech voice output devices*", ITU-T Recommendation P.85, International Telecommunication Union, June, 1994.