The effect of rhythm unit length on the duration of vowels in Serbian

Maja Marković and Tanja Milićev
University of Novi Sad, English Department
majamarkovic@ff.uns.ac.rs and *tanja.milicev@ff.uns.ac.rs*

Abstract: The paper presents the results of two studies of vowel duration depending on the length of rhythm unit in standard Serbian. The goal of the studies was to find out whether the number of unstressed syllables in one rhythm unit affects the duration of stressed vowels in Serbian. The results of both studies reveal a tendency towards the reduction of the duration of stressed vowels with the expansion of a rhythm unit. The studies were carried out in order to determine the nature of rhythm in Serbian and establish its typology (stress-timed or syllable-timed). The general purpose of the research was to establish whether vocalic length in relation to the rhythm unit length can be applied in speech synthesis to improve the performance of the existing speech synthesizer for the Serbian language.
Key words*:* rhythm unit, stressed vowel, stress-timed , syllable-timed languages, vocalic duration

1. Introduction

The prosodic domain investigated in this study is the temporal dimension of speech, which could be subsumed under the cover term 'speech rhythm'. In particular, it presents the results of an experimental investigation of the effects of the number of syllables in a rhythm unit on the duration of stressed vowels in Serbian. The central question addressed in our study is the interaction between some of the important factors which determine segmental duration, such as the position of the stressed word in an utterance, inherent vocalic duration arising from the articulation type of each vowel and the number of stressed and unstressed syllables within a rhythm unit (foot). We ultimately attempt to address the question of prosodic typology and find out whether Serbian shares the characteristics with the languages traditionally referred to as stress-timed or syllable-timed.

While the effect of the above-mentioned prosodic categories on rhythm has been extensively studied for some languages, there is very little knowledge on prosodic characteristics of Serbian, especially supported by experimental data. Our goal was therefore both theoretical and pragmatic: apart from illuminating the popular contrast between stress and syllable isochrony, the practical motivation for the investigation was to find out whether the temporal characteristics of segments could be applied to improve the performance of the Text-To-Speech (TTS) synthesizer developed for the Serbian language.

2. Rhythm, needs for studying it, approaches

The traditional division of languages between stress-timed and syllable-timed, proposed by Kenneth L. Pike (1945), dominated in the phonological theory and textbooks, as well as in the phonetic training classrooms for almost half century. According to this dichotomy, certain languages, such as English or Russian, are characterized by stress isochrony, the phenomenon that stresses tend to occur at approximately even time intervals. The consequence of accentual isochrony is that unstressed syllables have to be reduced, and the more of them, the more contracted they'll be, in order to successfully be 'squeezed' into rhythmic groups. In syllable-timed languages, on the other hand, each syllable tends to take

up an approximately equal amount of time, thus creating the impression of syllabic isochrony. The classical example of a syllable timed language is French.

However, with the advent of experimental acoustic analysis, it was found that there is no conclusive evidence to support either stress or syllable isochrony. The research in the field proved that stress isochrony is more or less a perceptual impression. The work of some linguists (Setter, 2006; Dellwo et al. 2004; Ordin and Setter 2008, etc.) suggests that certain characteristics are shared by the languages which were traditionally regarded as stress-timed. This is above all the difference between stressed, unstressed and weak syllables, which is not to be found in the languages labeled as 'syllable-timed'. The question of vowel/syllable reduction seems to play the most significant role when assigning a language to either the stress-timed or the syllable-timed category. The research of the above-mentioned linguists involves the instrumental measurements of syllable and segmental duration by which a language is assigned to one of the categories. The parameters calculated, among others, are: speech rate, mean syllable duration, mean duration of consonantal intervals, mean duration of vocalic intervals, standard deviations of vowel, consonant and syllable length, percentage of vocalic intervals etc.[1]

Other studies (Dauer 1983; Arvaniti 2009) show that rhythm should not be reduced to timing and that equating the two has led to circularity and a psychologically questionable conceptualization of rhythm in speech. Instead, the authors propose that research on rhythm be based on the same principles for all languages, and not the widely accepted division of languages into stress- and syllable-timed. Moreover, as has been shown in more recent studies (e.g. Arvaniti 2009), the notion of rhythmic types is problematic and largely incompatible with psychological evidence on rhythm.

2.1. Need for this study

In addition to theoretical concerns, our study is also motivated by the need to solve the problems of temporal relations in the Text-to-Speech Synthesizer (TTS System) for the Serbian language, developed at the Faculty of Engineering at the University of Novi Sad. Despite the high quality of the synthesized speech, its perception is currently hindered by inappropriate segmental durations, especially in terms of vocalic length. The duration of segments in the TTS is assigned regardless of the phonological and metric contexts in which the segment occurs. Namely, a vowel is assigned a fixed length value, generally on the basis of the division of vowels into four categories: 1) long stressed; 2) long unstressed; 3) short stressed; and 4) short unstressed. This leads to inappropriate and unnatural auditory effect.

3. Research: methodology, subjects, corpus

There are several reasons why the unit taken for the analysis is the vowel, and not the syllable. Being the carriers of the prosodic characteristics, vowels are prosodically the most salient elements, with the greatest acoustic output. Their temporal characteristics are perceptually dominant. In addition, vowels have proved to represent the smallest prosodic domain (Van Heuven 1994). Also, in comparison to vowels, consonants exhibit a lesser degree of variation, depending both on the grammatical role – pragmatic or syntactic (broad or narrow focus) or the speech tempo. Our final reason is drawn from phonological theory, which has proved the existence of the hierarchical organization of the syllable, with the

[1] The most widely accepted measurements for quantifying rhythm are the percentage of vocalic intervals in speech and the standard deviation of consonantal duration (%V and ΔC), proposed by Ramus et al. (1999), and the pairwise variability indices nPVI and rPVI (pairwise comparisons of successive vocalic and intervocalic, or consonantal, intervals) introduced by Grabe and Low (2002).

vowel as its central element and the asymmetrical relation between the syllable onset and the coda, where the two do not equally contribute to the syllable weight.

3.1. The pilot study

The first part of the research was a pilot study, where two male speakers were recorded reading words and phrases in isolation (115 in total). Each word/phrase contained a single stress on the first syllable. The phrases were presented to the subjects on a laptop screen one by one, each subsequent phrase containing one more unstressed syllable (1).

(1) a. **pút** 'road'
b. **kúpi**
buy.IMP.SG
'buy'
c. **kúpiću**
buy-will.AUX.1SG.CL
'I will buy'
d. **kúpićemo**
buy-will.AUX.2PL.CL
'We will buy'
e. **kúpićemo ga**
buy-will.AUX.2PL.CL it.ACC.CL
'We will buy it'
f. **kúpićemo mu ga**
buy-will.AUX.2PL.CL him.DAT.CL it.ACC.CL
'We will buy it for him'

In each of the phrases stress remained on the first syllable of the lexical word. A number of syllables added were either new bound morphemes (usually inflectional suffixes) or clitics, adjoined accentually to the existing lexical word.[2]

The subjects were allowed enough time between each subsequent phrase, and were instructed to pronounce the words or phrases as naturally as possible, despite the fact that they were out of a natural sentence context (although they were meaningful parts of potential sentences).

The speakers were recorded in the sound proof booth of the Faculty of Philosophy, Novi Sad University, directly on PC in *wav* file format, at 44.1 kHz sampling rate. The recorded material was acoustically analyzed using the Praat software for speech analysis. The results were statistically analyzed, providing the data on the mean duration of stressed vowels and the ratio between the longest and shortest durations of each stressed vowel.

In the choice of material we paid particular attention to cover all vocalic possibilities of the Serbian language. Serbian has five vowels /a, e, i, o, u/, but each of these can be in its short or long as well as falling or rising realizations (traditionally referred to as short falling, short rising, long falling and long rising accents), which makes the total of 20 vocalic possibilities. The phonotactic rules of Serbian excluded the possibility of rising vowels (either short or long) in monosyllables. Special attention was paid to another important factor, i.e. the consonantal environment. All vowels were preceded and, more importantly, followed by voiceless obstruents, for two reasons. The voiceless environment was chosen primarily because the voicing value of consonants after the vowel greatly affects the

[2] In terms of Selkirk (1984), the unit analyzed is prosodic (or phonological) word, which is also the domain of clitic attachment.

duration of the vowel. Another reason was the ease of analysis, as the fundamental frequency (F0) of voiced consonants could also affect the placement of boundaries between segments.

The results of the pilot study showed that by expanding the rhythm unit by adding unstressed syllables, the duration of the stressed syllable, all other things being equal, showed the decreasing tendency. The tendency was the most conspicuous with long vowels, but throughout the recorded material the mean value of the decrease was over 20% from monosyllabic feet to 5 or 6 syllable feet. The calculation of average duration values reveals that long vowels in mono or disyllabic feet are on average 41% longer than in 5 or 6 syllable feet. Short vowels are on average 24% longer in mono or disyllabic feet than in 5 or 6 syllable feet. According to some previous findings in the literature, the threshold for perceiving the difference in duration is 10% (Bakran 1996) or 5-15% (Nooteboom 1997), which means that the decrease in the duration found in our study was perceptually significant.

3.2. The second study

Since the material of our pilot study was rather artificial, we created a set of new sentences which were again very carefully chosen regarding various criteria. The sentences were the answers to a set of corresponding questions.

Each sentence contained two stressed words, since one of the goals was to analyze whether the position in the sentence also affected the duration of the vowel. The sentences contained two relatively symmetrical elements of the same number of syllables. Another thing we took care of was to achieve unmarked focus, i.e. to avoid contrastive or narrow focus of any kind. Here we encountered the difficulty of finding monosyllables in the sentence final position (unmarked word order), especially due to the rich morphology of Serbian. Another restriction imposed again involved the phonetic requirements, i.e. the vowels were placed in voiceless consonantal environment. The last criterion we set was to only analyze long vowels, as we expected they would exhibit a greater effect of the decrease in duration. The direction of the tone, falling or rising, was not taken into account, as it proved not to significantly affect the duration of vowels.

The subjects of this study were 4 first year students of English at the Faculty of Philosophy, 1 male and 3 female. Each of them speaks standard Serbian as their first language. The students were first familiarized both with the questions and the answers, but the purpose of recording was not explained. They were instructed to read the sentences as naturally as possible. One of the students asked a question, and the other one, seated by the microphone, was recorded while reading the answer. The recording was made as a digital *wav* file, sampled at 44.1 kHz frequency. The recorded material was analyzed using the Praat software for speech analysis and the duration of stressed vowels was measured.

The results of measurement, expressed as mean vocalic durations, are presented in Tables 1 and 2. Table 1 shows the mean duration of the vowel in the first stressed word in the sentences. Table 2 shows the mean duration of the vowel in the second stressed word. Sign ' ˊ ' indicates a stressed syllable and sign ' ˇ ' an unstressed syllable in a foot. All the examples analyzed had initial stress. (2) illustrates a set of examples in which the duration of /a/ was measured in the stressed syllable.

(2) a. **sát** 'hour'
b. **sáti** 'hours'
c. **sátima**
hours.INST
d. **sátima ga**

hours.INST him.ACC.CL
e. **sátima smo ga**
hours.INST be.AUX.PRES.1PL.CL him.ACC.CL
f. **sátima ćemo ga**
hours.INST will.AUX.2PL.CL him.ACC.CL

Table 1. Mean vocalic duration (in msec) of the vowel of the first stressed word.

Stress pattern	i	e	a	o	u
´	133	128	157	156	142
´ ˘	114	129	131	160	116
´ ˘ ˘	99	118	130	139	103
´ ˘ ˘ ˘	94	111	127	120	97
´ ˘ ˘ ˘ ˘	86	n/a	121	n/a	96

Table 2. Mean vocalic duration (in msec) of the vowel of the second stressed word

Stress pattern	i	e	a	o	u
´	n/a	n/a	n/a	n/a	n/a
´ ˘	129	150	174	164	140
´ ˘ ˘	100	121	158	132	114
´ ˘ ˘ ˘	82	125	147	128	110
´ ˘ ˘ ˘ ˘	n/a	n/a	130	n/a	103

The results show several significant tendencies regarding vocalic duration in sentences:

(a) The stressed vowels of the last accented word in the utterances examined clearly and consistently tend to be longer than the stressed vowels of the first word. This is hardly surprising, considering the fact that clause final position is a typical focus position in Serbian. In addition, clause final element is (also) marked by pre-boundary lengthening. In order to achieve natural sounding speech, a TTS should undoubtedly aim at implementing rules for assigning additional length value in this position/within this prosodic domain.
(b) Inherent vowel length, depending on the quality of the vowel, i.e. the degree of aperture used for the articulation, is an important factor at the sentence level as well. As has been known from earlier research, vowels produced with a more open position of the jaw/tongue, such as /a/, tend to have longer duration than the high vowels /i, u/ (Lehiste and Peterson 1961, Bakran 1996, Kent and Read 2002; similar findings for the Serbian language in Marković and Bjelaković 2008). This is also clearly shown in this study (Tables 1 and 2). The low vowel /a/ is consistently the longest, whereas the high vowels /i, u/ are the shortest. The other vowels generally comply with the expected tendencies. The only exception is the relatively longer than expected realization of /o/. While in the sentence final position (Table 2) its duration is more or less as expected, in the sentence initial position, it is significantly longer, approximating the duration of /a/ (Table 1). This remains an open question, which will be addressed in subsequent research.

(c) All the mean duration values presented in Tables 1 and 2 undeniably prove that the expansion of a rhythm unit by adding unstressed syllables contributes to the reduction of length of the vowel in the stressed syllable. The differences in duration of stressed vowel in a single syllable foot and a five syllable foot range between 15 and 59%, depending on the vowel and the positions in the utterance. If we recall that the duration difference of 10% is perceptually significant (cf. the discussion in Section 3.1) it can be assumed that this temporal characteristic is crucial for natural sounding speech, and should therefore not be neglected in speech synthesizing systems.

4. Conclusion

The study presented in this paper is an attempt of tackling some of the elusive characteristics of spoken language, i.e. the temporal characteristics of segments or, more specifically, the duration of vowels in stressed syllables in Serbian. Being carried out under carefully controlled conditions, the results can be said to reveal but the tip of an iceberg when it comes to the problem of segmental duration in longer stretches of speech. However, we believe that controlled experiments are a good start which gives a sound basis for further investigations.

The two experiments carried out clearly indicate that Serbian shares some of the characteristics of the so-called 'stress-timed' languages, with reduced unstressed vowels and adjusted vocalic length in longer rhythm units. In this respect, Serbian exhibits the so-called 'compensatory shortening', the phonological phenomenon by which, stressed syllables of a foot in certain languages tend to be compressed depending on the number of the following unstressed syllables of that foot or phonological word. Such a tendency, also called intersyllabic compensation, according to the results of previous studies, seems to be a characteristic of stress-timed languages (Bertinetto 1989).

However, the study strongly suggests that focusing on durational measurements alone is not sufficient, as local durational contrasts clearly depend on the metric characteristics of speech. In this sense, the results of this research are in line with the previous studies which show that the meter is the important domain of the mental representation of rhythm (Lerdahl and Jackendoff 1983).

Obviously, the traditional dichotomy between stress and syllabic isochrony is oversimplified, while various factors have significant roles in assigning duration to segments in natural speech. The ones studied in this research, i.e. the number of syllables in a foot and the position of stress within an utterance, play a relevant part in prosodic characteristics of spoken language. The results of the study clearly show that the duration of vowels directly depends on the number of syllables in the foot, as well as on the position of the stressed word in an utterance. Other factors (inherent vowel length, different phonemic context (especially voiced/voiceless environment) are also relevant for assigning segment duration, but they do not appear to be as significant as the two factors studied in this research.

As regards the advancement of the TTS system for the Serbian language, the research also opens up a number of questions and fields of future analysis. Firstly, it suggests a need for a range of perceptual experiments to check the importance of these findings, and, secondly, it imposes the task of investigating how these results interact with the other factors influencing the temporal characteristics of speech.

References

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm**.** *Phonetica* 66: 46-63.
Bakran, J. (1996). *Zvučna slika hrvatskoga govora*. Zagreb: Ibis grafika.

Bertinetto, P. M. (1989). Reflections on the Dichotomy 'Stress' vs. 'Syllable-timing'. Revue de Phonétique Appliquée, Mons, 99-130.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics 11*: 51– 62.

Dellwo et al. (2004). Bonntempo-corpus and bonntempo-tools: a database for the study of speech rhythm and rate. *INTERSPEECH-2004*. 777-780.

Grabe, E. and Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.) *Papers in Laboratory Phonology*, Volume 7. Berlin: Mouton de Gruyter, 377-401.

Heuven, V.J. Van. (1994). What is the smallest prosodic domain? In P. Keating (ed), *Papers in Laboratory Phonology III: phonological structure and phonetic form*. Cambridge: Cambridge University Press, 76-98.

Kent, R. D. and C. Read. (2002). *Acoustic Analysis of Speech*. 2nd edition. Singular: Thomson Learning.

Lehiste, I. and Peterson G. E., (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America* 33: 419-425.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press, Cambridge, MA.

Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'syllable-timing'. *Singapore English*. *Language and Speech* 43: 377–401.

Marković, M. and Bjelaković I. (2008). Kvantitativno-kvalitativni odnosi akcentovanih vokala u govoru Novog Sada. Zbornik radova *Digitalna obrada govora i slike, Kelebija 2008*, 25-29.

Nooteboom, S. (1997). The prosody of speech: melody and rhythm. In W.Hardcastle and J.Laver (eds), *The Handbook of Phonetic Sciences*. Blackwell, 640-673.

Ordin, M. Y. and Setter, J. (2008). Comparative research of temporal organization of the syllable structure in Hong Kong English, Russian English, and British English. *Proceedings of XX Session of the Russian Acoustical Society*, Moscow, 653-656.

Pike, K. (1945). *Intonation of American English*. Ann Arbor: University of Michigan Press.

Ramus, F. Nespor M. and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 72: 265-292.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Setter, J. (2006). Speech rhythm in World Englishes: The case of Hong Kong. TESOL Quarterly 40/4: 763-782.