Effects of lexical stress and speech rate on the quantity and quality of Slovak vowels

Štefan Beňuš [1,2], *Katalin Mády* [3]

[1]Dep. of English and American Studies, Constantine the Philosopher University, Nitra, Slovakia
[2]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[3]Institute of Phonetics and Speech Processing, University of Munich, Germany

sbenus@ukf.sk, mady@phonetik.uni-muenchen.de

Abstract

We investigate the relationship between vowel quantity and the utilization of formant space in Slovak, and how prosodic variation in speech rate and lexical stress marking affects this relationship. Slovak presents a common five-vowel system with full phonemic quantity contrast for all vowels in all positions. We found that 1) phonemic quantity contrast in Slovak is salient and minimally affected by lexical stress and speech rate, and 2) shortening due to phonemic contrast and destressing but not due to speech rate, are accompanied by vowel space contraction. We compare the results to the geographically neighboring languages Czech and Hungarian that display similar prosodic characteristics to Slovak.

Index Terms: vowel quantity, vowel quality, lexical stress, speech rate, Slovak

1. Introduction

Vowel duration is one of the major features used in signaling prosodic distinctions such as the presence and absence of lexical stress, prosodic boundaries, or phrasal accents. The system underlying segmental quantity, however, is domain-specific and language-specific, and thus relevant for the cognitive system that underlies speech. For example, the articulatory strategies for achieving acoustic lengthening in accented syllables are different from the strategies used for pre-boundary lengthening or speech tempo variations [1]. In languages that use vowel duration for signaling phonemic contrasts, the competing requirements of the segmental and prosodic systems are resolved in a language specific manner. For example, in Finnish, utterance-final lengthening is regulated to preserve the phonemic quantity distinctions, and although final lengthening is a universal tendency, its implementation is proposed to be language-specific [2].

In addition to prosody, vowel quantity also interacts in language-specific ways with vowel quality. For example, Hungarian has 14 monophthongs that can be arranged in 7 quantity pairs based on their phonological behavior, but vowel quality diverges largely for two of these pairs [3]. It was observed in [4] that the phonetic realization of the quantity contrast correlates with the realization of the quality contrast: vowels with robust quality distinction, such as /a/ or /e/, display robust quantity contrasts, while overlapping durations have been observed for vowels with minimal quality differences such as /i/, /y/, and /u/. The opposite tendency characterizes the interaction of quality and quantity distinction in Czech [5]. At least one out of five Czech vowels with phonemic quantity contrasts, namely /i/, is also distinguished by quality: long [i:] is tense while short [ɪ] is lax. Contrary to Hungarian, qualitatively different [i:] and [ɪ] in Czech are produced with considerable durational overlap, whereas vowel pairs without the quality distinction are consistently distinguished durationally.

Slovak is a West-Slavic language genetically closely related to Czech and presents a five-vowel system of monophthongs /i/, /e/, /a/, /o/, /u/ with full phonemic quantity contrast for all vowels in all positions. Although Hungarian is Finno-Ugric and thus genetically unrelated to Slavic, all three languages are related geographically through extensive language contact and display remarkable prosodic similarities by fixed left-most lexical stress and phonemic vowel quantity distinctions. Standard literature on Slovak [6] claims that quantity contrast does not affect the quality of vowels apart from long /a:/ that is said to be slightly more open than short /a/. This claim corresponds to the intuitions of native speakers, but careful quantitative investigation of Slovak vowel productions is missing.

The primary focus of the present study is to investigate the relationship between vowel quantity and the utilization of formant space in Slovak, and how prosodic variation in speech rate and lexical stress marking affects this relationship. More specifically, there are potentially three sources of formant space contraction related to quantity: shortening due to the phonemic quantity contrast, due to absence of lexical stress, and due to fast speech rate. We investigate the degree of quantity variation of these three sources and to what extent they are accompanied by centralization. By including lexical stress and tempo variation, our study also fills the gaps in the studies of Czech and Hungarian.

2. Methodology

Primary data for the study come from the recordings of two native Slovak speakers (one male and one female) reading prompt sentences in two speech rates. The stimuli consist of nonsense words in the forms C_iVC_ia and C_iaC_iV embedded in a prompt sentence *Hovoríme ___ podobne a ___ podobne* 'We say ___ similarly and ___ similarly'. Symmetrical voiceless plosives /p/, /t/, /k/ flanked the target vowels from both sides, and vowel targets included all Slovak monophthongs /i/, /e/, /a/, /o/, /u/, /i:/, /e:/, /a:/, /o:/, /u:/.

In order to avoid somewhat unnatural methods for achieving speech rate variation such as metronome prompting, a combination of approaches in [7] and [8] was used. First, during a pre-recorded session we asked speakers to read prompt sentences in their natural tempo and assigned a number 10 to this tempo. Then we practiced reading in twice that rate (20) and half that rate (5). Additionally, the rates of 15 and 40 were practiced. When subjects were comfortable with the rate variation, we recorded a subset of the test sentences in blocks for each of the five rates. From this subset we selected two rates (normal and fast), and example sentences from those rates, that yielded good overlap in the

durations of short vowels in the normal rate and long vowels in the faster rate. The example sentences were then presented auditorily and randomly before each 4th - 8th prompt sentence during the actual recording session, and the subjects were instructed to match the rate of the test sentences as closely as possible to the rates of the example sentences. The recording sessions alternated blocks of normal and fast speech rates.

This procedure resulted in 240 target words for each subject (3 consonants, 10 vowels, 2 positions, 2 rates, and 2 repetitions), which gives 528 tokens of target vowels (including both [a] vowels that occurred in words *papa*, *tata*, and *kaka*). Subjects were also instructed, and reminded if necessary, not to make any prosodic breaks during the test sentences. Data were recorded using the SpeechRecorder interface [9] with a head-mounted condenser microphone and digitized at 44.1 kHz.

Automatic SAMPA annotations for the target sentences were created and then forced aligned to the acoustic signal using the HTK toolkit. The phoneme boundaries were hand-corrected using Praat [10], and annotations for syllables and words based on the phoneme alignment were added. The boundaries between consonants and vowels were marked at the beginning of the modal voice on the left of the vowel, and the cessation of formants on the right. A Praat script was used to automatically extract the duration of each vowel, and first and second formants as median values from the 15 ms window around the temporal midpoint of the vowel. All formant values were converted to Bark scale [11] to better describe perceptual characteristics of quality variation.

To assess the variation in the utilization of the vowel space, we computed the centroids of the vowel space for each speaker from all vowels produced by that speaker as the F1 and F2 means, and the Euclidean distances for each F1-F2 data point from the respective centroid of that speaker.

3. Results

3.1. Prosody and vowel quantity

We start with a descriptive analysis of vowel durations in relation to the three factors: phonemic contrast, presence of lexical stress, and speech rate. Figure 1 illustrates the effect of lexical stress on vowel durations in the fast (top) and normal (bottom) speech rate. The box plots show that long vowels are significantly longer than short vowels irrespective of speech rate and the presence of lexical stress. Independently of the prosodic factors, subject 2 marked the phonemic contrast with greater ratio of long/short vowels, and the durational contrast of the vowel [a] was in general greater than with the other four vowels for both subjects. A tendency for the weakening of the quantity contrast could be observed for vowels [i] and [e] for subject 1 especially in the unstressed position in both rates, and also for [u] in the fast rate. Prosodic weakening due to the absence of stress and fast speech rate affected long vowels more than short vowels. Similar results were obtained for the duration of the CV syllables in which these vowels occurred.

The information in Figure 1 can be numerically expressed also as ratios for pairs of long/short vowels and stressed/unstressed vowels in each condition. These ratios show that the characterization of Slovak long vowels as twice as long as the short ones is supported in our data because all long/short ratios reached above 1.5 and on average they were around 2. A minor but consistent compression of the quantity contrast was observed in the fast rate ($\Delta \approx 0.3$ in all conditions). The long/short ratio differed robustly from the stressed/unstressed ratio for subject 2 (with means of 2.1 for length and 1.1 for stress) but the ratios for [e] and [u] for subject 1 in fast rate overlapped. Hence, for this subject, the short-long contrast can be realized similarly to the stressed-unstressed contrast for some vowels. Subject 1 also produced consistently greater quantity differences between stressed and unstressed vowels than subject 2.

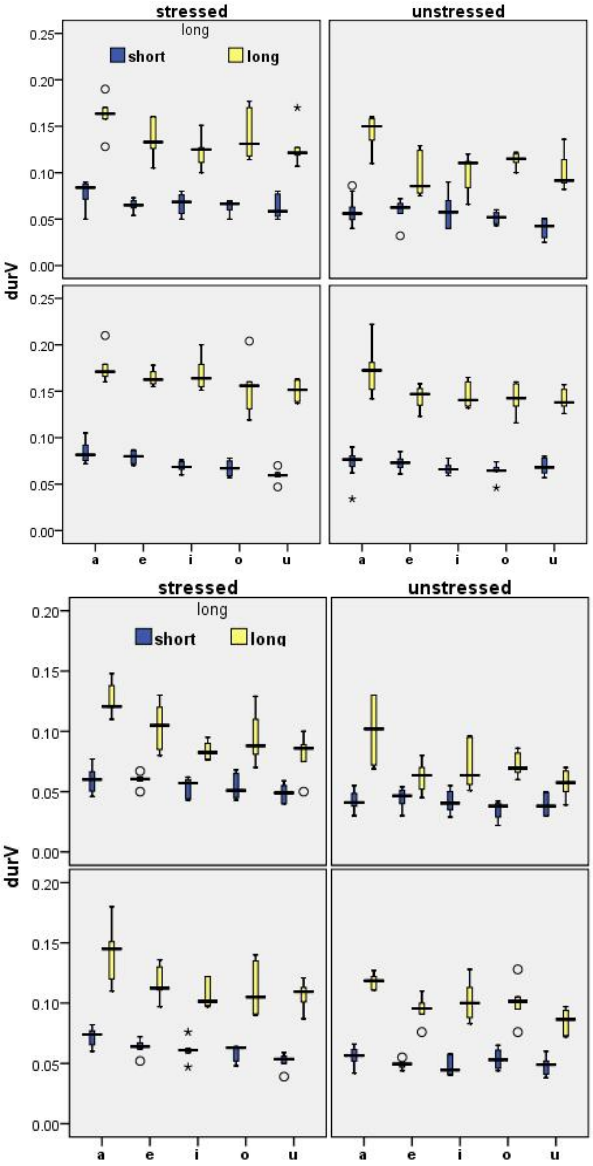

Figure 1: *Vowel durations (in sec.) separately for 5 vowels clustered by phonemic length, paneled vertically by subject and horizontally by lexical stress; fast rate on the top and normal rate on the bottom.*

3.2. Prosody and vowel quality

Similarly to duration, we analyze the formant space characteristics related to lexical stress and speech rate descriptively. First, pooling all vowels together and assessing the compression of the vowel space with Euclidean distances, the boxplots in Figure 2 show that centralization of short vowels compared to the long ones is the most consistent trend for both subjects and speech rates. Centralization of unstressed vowels compared to the stressed ones can also be observed for both rates in subject 1 and to a small degree also in fast rate in subject 2. Fast speech rate does not seem to be realized with

more centralized vowels than normal rate. Hence, speech rate caused greater effects in quantity than lexical stress, while the reverse is true for vowel quality.

The plots in Figure 3 provide information about formant space utilization for individual vowels. It can be observed that with few exceptions short vowels are generally more centralized compared to the long ones in all prosodic positions. Looking at the normal speech rate we see that short vowels tend to centralize due to the absence of stress much more than the long vowels. Comparing the effects of lexical stress (horizontally adjacent plots in columns 1-2 and 3-4) with speech rate (columns 1-3 and 2-4) we observe that the absence of lexical stress seems to lead to greater neutralization than fast speech rate. Both of these observations confirm the conclusions drawn from Figure 2. Interestingly, front vowels, and especially [e], tend to be centralized the most, then [a], and back vowels centralize the least. This difference between front and back vowels is particularly robust in the unstressed positions. Recall that the quantity differences caused by lexical stress especially for short vowels were minimal, yet the quality differences were robust. The opposite applied to speech rate variation: it produced robust quantity differences particularly for long vowels but the quality differences were less robust. Finally, Figure 3 shows that despite pervasive contraction of vowel space, the five phonemic vowel qualities remain remarkably well separated in all conditions with the exception of some overlaps in [u] and [o].

Analyzing the effect of vowel centralization as a function of vowel duration, Figure 4 displays Euclidean distance plotted against vowel duration. We pooled together speech rate and phonemic length conditions. Considering both the slopes of the regression lines and R^2 coefficients we observe that /i/ displays positive slopes and high coefficients consistently for both subjects and stress conditions. Vowel /a/ shows high correlations in three out of four conditions (except unstressed, subject 1), and vowel /e/ behaved similarly to /i/ for subject 1 in both conditions. Finally, vowel /u/ reached a high R^2 value in the unstressed syllables for subject 2.

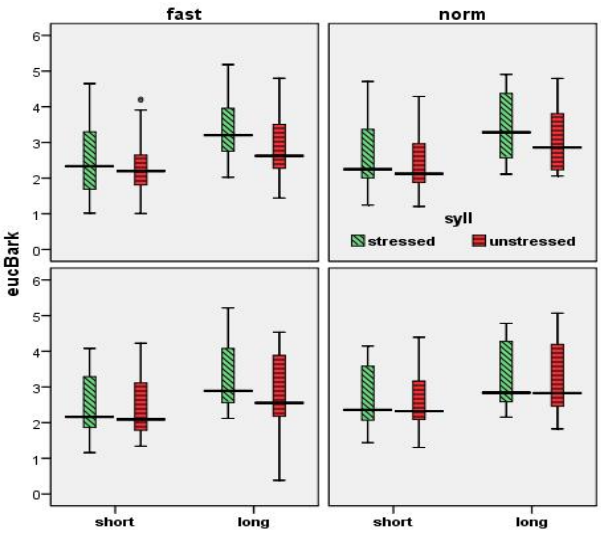

Figure 2: *Euclidean distance in short and long vowels clustered by lexical stress, paneled vertically by subject and horizontally by rate.*

4. Discussion and conclusions

The phonemic contrast in vowel quantity in Slovak was robustly present in our data. While prosodic conditions of lexical stress placement and speech rate variation affected particularly long vowel durations to a great extent, the durations of short and long vowels overlapped minimally within prosodic conditions. Phonemic contrast and speech rate produced robust effects on quantity, while lexical stress yielded smaller effects. A tendency for the neutralization of the quantity contrast was observed for front vowels in unstressed position.

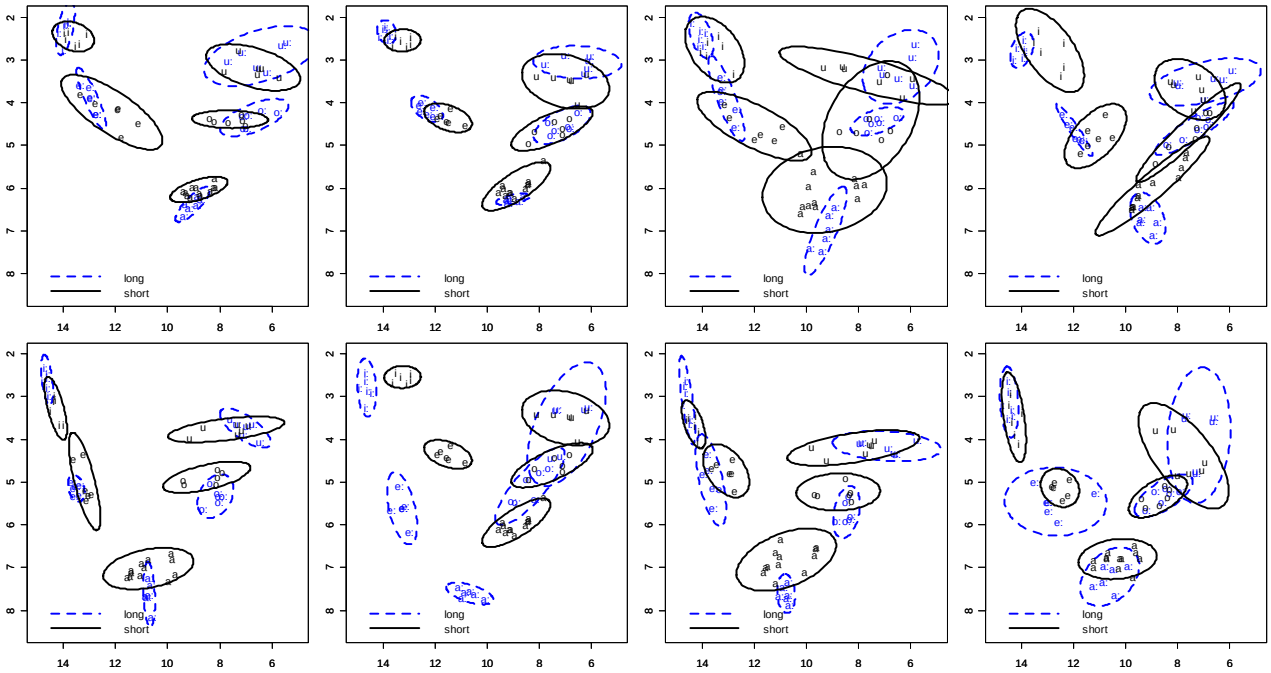

Figure 1: *Effects of stress and rate on the utilization of formant space (x = F2, y = F1, both in Bark) for short and long vowels. From left to right: normal rate stressed syllables, normal rate unstressed syllables, fast rate stressed syllables, fast rate unstressed syllables; subject 1 on the top and subject 2 on the bottom.*

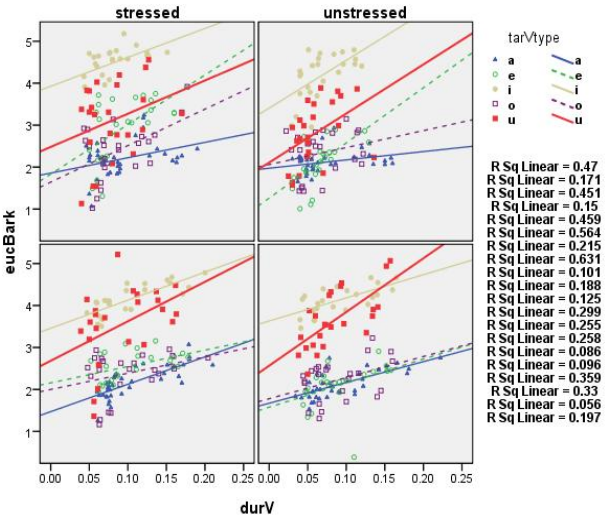

Figure 4: *Euclidean distance (Bark) plotted as a function of vowel duration pooling together long and short vowels and both speech rates, Subject 1 in the top two plots and Subject 2 in the bottom two. The R^2 lin. coefficients are ordered in four groups of five {a, e, i, o, u} counterclockwise from the bottom-right.*

In terms of quality, phonemic contrast and lexical stress produced greater effects in the contraction of the vowel space than speech rate variation. Positive correlations between vowel quantity (both in the continuous and discrete dimensions) and centralization were observed for /i/, /e/ and /a/. Front vowels /i/ and /e/ tended to be centralized in the horizontal (F2) dimension while /a/ showed the greatest effects in the vertical (F1) dimension. Hence, quality effects were observed for /a/ as the vowel with the most robust quantity contrast as well as for vowels /i/ and /e/ that were produced with less robust quantity contrasts. In this sense, phonemic vowel contrasts in Slovak are phonetically neutralized in prosodically weak positions more in terms of quality than quantity, yet none of the neutralizations disturb the phonological system. Slovak thus differs both from Czech, where additional quality differences lead to less consistent durational distinctions, and from Hungarian, where quantity and quality distinctions correlate.

We also observed that speech rate affected durations robustly but quality only minimally, while phonemic length and stress tended to affect both quantity and quality. Moreover, the quality effects were observed in both the horizontal (F2) and vertical (F1) dimensions. It has been argued that phonetic vowel neutralizations in the horizontal dimension are attributable to the target undershoot resulting from greater consonantal overlap while neutralizations in the vertical dimensions arise from the adjustments to the spatio-temporal characteristics of the jaw movements related to the phonological notion of sonority ([1]). Our data suggest that neutralizations caused by prosodic weakening in unstressed short vowels have access to both of these strategies and the coupling of shorter durations with smaller displacement generally support the task-dynamics model of Articulatory Phonology, e.g. [12]. However, prosodic weakening resulting from faster speech rate in our data seems to be fundamentally different in that it affects durations more than quality and is consistent with the strategy of increasing gestural velocity without significant adjustments to the spatial vocalic targets or consonant-vowel overlaps. It was argued [7] that speech gestures in fast rates involve simplified pre-programmed movements whereas slow speech gestures consist of multiple sub-movements that may be influenced by feedback mechanisms. Our data fit in this general paradigm if we assume that lexical stress, the quantity and quality of vowels are all phonological attributes with interacting articulatory requirements, and they result in mutually interdependent articulatory submovements (e.g. through gestural blending [12]). Speech rate variation, on the other hand, as a non-phonological dimension, primarily affects the stiffness of the simplified unitary movements in the sense of [7].

Finally, asymmetries between quality and quantity marking of vowel contrasts might give rise to socio-linguistic sound changes. Vowel systems of regional varieties of Hungarian differ with regard to length: in some western dialects there is no quantity distinction for high vowels and since they are not supported by quality distinctions, the contrast is being neutralized. In Czech, [5] observed that western speakers (Bohemians) preferred quality cues for the differentiation of [i:] - [ɪ] contrast, while eastern speakers (Moravians) tended to rely more on durational cues. Since Slovakia borders with the east of the Czech Republic, and we found that durational cues are robustly present, our data support the suggestion of [5] that the difference between the western and eastern Czech speakers could be explained by sound change in progress.

5. Acknowledgements

This work was supported in part by the Alexander von Humboldt Foundation, the Slovak Ministry of Education grant KEGA 3/6399/08, the Slovak Research and Development Agency project APVV-0369-07, and the German Research Council excellence project "Speaker characteristics and language society".

6. References

[1] Beckman, M.E., Edwards, J., and Fletcher, J. “Prosodic structure and tempo in a sonority model of articulatory dynamics”, In: Laboratory Phonology II, 68-86, 1992.

[2] Nakai, S., Kunnarib, S., Turk, A., Suomi, K., Ylitalo, R., “Utterance-final lengthening and quantity in Northern Finnish”, Journal of Phonetics 37(1), 29-45, 2009.

[3] Siptár, P., Törkenczy, M., “The phonology of Hungarian”, Oxford, University Press, 2000.

[4] Mády, K. and Reichel, U.D., “Quantity distinction in the Hungarian vowel system--just theory or also reality?”, Proc. 15th ICPhS, 2007.

[5] Podlipský,V. J., Skarnitzl, R., and Volín, J., “High Front Vowels in Czech: a Contrast in Quantity or Quality, Proceedings of Interspeech, 132-135, 2009.

[6] Kráľ, A. and Sabol, J., Fonetika a fonológia. Bratislava: Slovenské pedagogické nakladateľstvo, 1989.

[7] Adams, S.G., Weismer, G. and Kent, R. D., “Speaking rate and speech movement velocity profiles”, Journal of Speech and Hearing Research 36, 41–54, 1993.

[8] Hoole, P. and Mooshamer, C., “Articulatory analysis of the German vowel system”, In: P. Auer, P. Gilles and H. Spiekermann (eds.), Silbenschnitt und Tonakzente. Niemeyer, Tübingen, 129-152, 2002.

[9] Draxler, Ch., and Jansch, K., SpeechRecorder: A Universal Platform Independent Multi-Channel Audio Recording Software,http://www.phonetik.uni-muenchen.de/Bas/software/

[10] Boersma, P, and Weenink, D., Praat: Doing phonetics by computer, http://www.praat.org, 2009.

[11] Traunmüller, H. “Analytical Expressions for the Tonotopic Sensory Scale”, J. Acoust. Soc. Am. 88, 97-100, 1990.

[12] Fowler, C. A., and Saltzman, E., “Coordination and Coarticulation in Speech Production”, Language and Speech, 36(2,3), 171-195, 1993.