



## Stress and phonemic length in the perception of Slovak vowels

Štefan Beňuš<sup>1</sup>, Katalin Mády<sup>2</sup>

<sup>1</sup>Constantine the Philosopher University, Nitra, Slovakia, Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia

<sup>2</sup>Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary

sbenus@ukf.sk, mady@nytud.hu

### Abstract

We investigate the perception of phonemic vowel quantity contrast and its relation to word stress and vowel quality in Slovak, and fill the gap of missing experimental perception data for this language. We observe that both prosodically-driven undershoot of unstressed vowels and the functional load affect the perception of quantity contrast. Vowel quality plays some role in quantity identification for high vowels (/u/ and unstressed /i/) but Slovak seems to retain a robust quantity contrast for all examined vowel qualities.

**Index Terms:** vowel duration, Slovak, perception, word stress

### 1. Introduction

Vowel duration is fruitful ground for prosody research since it marks a wide array of linguistic and paralinguistic functions from phonemic contrast to prosodic phrasing and accenting, rhythm, and to emotionally-loaded dimensions like speech rate. Due to this complexity, understanding the relationships among these functions and their modeling in applications using speech still represent a challenge for current research. In this paper we follow up on the production study of the interactions of phonemic length, speech rate, and words stress in Slovak vowels [1] and explore the perception of phonemic quantity contrast and its relation to word stress and vowel quality in the same language.

Slovak is a West-Slavic language that has a phonemic distinction of vowel quantity in all five major vowel qualities [i, e, a, o, u], and also in two syllabic liquids [r, l]. Slovak has a left-most fixed primary word stress and relatively non-salient secondary stress on odd-number non-final syllables counting from the left. Quantity marking is not restricted by stress since both long and short vowels appear in both stressed and unstressed syllables. Despite this, long vowels tend to appear more frequently in unstressed vowels than in the stressed ones: [2] report that the ratio of long over short vowels in the first syllable is 0.15 and then it gradually increases: 0.26 in the 2<sup>nd</sup> syllable, 0.3 in the 3<sup>rd</sup>, 0.3 in the 4<sup>th</sup>, and 0.36 in the 5<sup>th</sup>. This suggests that non-initial syllables might be more functionally loaded for the quantity contrast than the initial ones, which is strongly linked to the rich derivational and inflectional morphology of Slovak and the presence of quantity marking in suffixes. Note, however, that syllable-related relative frequency of long vowels is taken here only as a 'raw/first' indicator of functional load in the absence of data for calculating it in a more standard way [3]. If the frequency data in [2] can be linked to functional load it would suggest that Slovak native speakers should be identifying short/long vowels in unstressed syllables better than in the stressed ones.

However, it is a common feature of languages that unstressed syllables tend to be produced with shorter duration and greater articulatory undershoot [4]. Impressionistically, Slovak has been described as a language with minimal differences between the quantity and quality of vowels depending on the presence of word stress. Only long /a:/ was described as reaching a lower jaw position compared to short /a/. Recent experimental studies on Slovak vowel quantity production [1,5] show that in laboratory speech, long vowels are about 1.5-2 times longer than short vowels but this ratio is likely to be lower in everyday spontaneous speech. The salient nature of quantity contrast in Slovak was illustrated also in [6] where 'mispronunciations' of native German speakers producing sentences in Slovak as their L2 were analyzed. 'Mistakes' in L2 Slovak productions in vowel quantity were the most frequently perceived problem. Additionally, [1] in a limited data from two subjects found that phonemic quantity contrast in Slovak is minimally affected by speech rate but also reported durational differences between stressed and unstressed vowels. Furthermore, experimental data on vowel production in the same study showed that shortening due to phonemic contrast and de-stressing were accompanied by vowel space contraction, but that the five vowel qualities were still well separated even in this contracted space.

Given that Slovak displays small quantity and quality differences depending on word stress, we would predict that the quantity distinction in unstressed vowels should be less robust than in the stressed vowels. Hence, the first goal of this paper is to examine whether the greater functional load of the unstressed syllables, or greater duration together with lower undershoot of the stressed syllables, play a greater role in the perception of the phonemic quantity contrast in Slovak.

Additionally, Slovak is geographically and historically closely linked to Czech, another West-Slavic language, and Hungarian, an unrelated Finno-Ugric language. These three languages share the robust presence of vowel quantity distinctions in their sound systems and the leftmost word stress pattern. In both Czech and Hungarian it has been recently observed that the quantity marking is becoming substituted by quality. In Hungarian, the relationship between quantity and quality depends on vowel height [7]. As has been shown by [8], this interdependency is also present in quantity perception. In Czech, the contrast between short and long /i/ is gradually becoming a contrast of quality rather than quantity both in production and perception [9]. It is then an open question if similar tendencies can be observed in Slovak.

The final issue to be investigated, which also relates to the cross-linguistic differences discussed above, concerns potential differences among Slovak vowels. For example, [2] suggested that /a/ is the only vowel with meaningful quality difference between long and short. [1] found that centralization affects all vowels, but /a/ might indeed be affected the most. Finally, [5] showed that the jaw

displacement in long and short vowels, which plays an important role in vowel duration and undershoot characteristics, also depends heavily on vowel height. Hence, it is plausible to expect that the differences in how the quantity contrasts are realized in individual vowel types in stressed/unstressed syllables might also be reflected in how their quantity contrast is perceived.

## 2. Methodology

Material for the perception experiment was created in the following way. First, we constructed near minimal pairs of words with the long and short target vowels [a a: o o: u u: i i:] in the first (stressed) and second (unstressed) syllable listed in Table 3. Long /e:/ is relatively restricted in marking lexical distinctions and was thus omitted from this study. Hence, 8 pairs of words were selected so that the target vowel was not word-final to prevent the effects of boundary lengthening, and was flanked mostly by coronal consonants (unstressed /u/ is the sole exception). Short and long words in each pair were impressionistically matched for word frequency and none of the words was extremely common or rare.

Table 3. Stimuli in IPA, target CVC sequences bolded

		Stressed	Unstressed
<i>i</i>	long short	<b>si:rovou</b> <b>sirotou</b>	<b>kuri:nou</b> <b>perinou</b>
<i>u</i>	long short	<b>su:di</b> <b>sudi</b>	<b>lopu:xom</b> <b>opuxom</b>
<i>o</i>	long short	<b>to:novi:</b> <b>tonove:</b>	<b>karto:nom</b> <b>antonom</b>
<i>a</i>	long short	<b>na:dobou</b> <b>nadanou</b>	<b>fasa:di</b> <b>osadi</b>

Target words were embedded in very similar meaningful sentences of low semantic weight so that none of the target words could be favored over the other. The target vowel quality did not appear elsewhere in the sentence. For stressed vowels, all material before the target sequence, and for unstressed after the sequence, were identical within the pair.

The sentence list was recorded by a single female speaker of standard Slovak multiple times instructed to match the prosody within each pair. The closest matching pair (auditorily and quantitatively) was then selected and the target CVC sequence was aligned to the signal. To obtain data about typical durations, we used a corpus of nonsense CVCa words from 5 speakers [5] and extracted min/max durations of long and short vowels in normal tempo (150 tokens for each vowel). Since the speaker for the prompt sentences was relatively fast, we decreased the max values by 60ms and obtained these ranges in ms: 50-220 for /i/, 40-190 for /u/, 50-200 for /o/, 40-260 for /a/. Target vowels were cut from the carrier sentence, and their duration was manipulated using Praat's PSOLA algorithm. The first and final 25% of the vowel duration were left unchanged to maintain formant transitions and the central 50% of the vowel was manipulated, creating 9 equal steps of quantity, of which only the first 7 were used to prevent unnaturally long vowels given the fast rate of the prompt sentences. Durations were kept identical across stressed and unstressed vowels. The procedure was applied to both originally short and long vowels, leaving their spectral characteristics unchanged. Thus, a vowel pair contained 7 duration-manipulated vowels with the spectral

cues of the originally short or long vowel. This procedure resulted in 112 manipulated vowels (4 vowel types, 2 quantities, 2 stress types, 7 durations). The manipulated vowels were embedded in the original sentences and the non-identical parts of the sentences were replaced by a waterfall noise matched for the intensity of the sentence.

In a forced-choice perception experiment, 17 subjects (12 F, 5M) aged 19-23 listened to 5 repetitions of each sentence (N = 560) in random order. Subjects were instructed to click on one of the two buttons; each button represented the sentence in Slovak orthography containing a short/long target vowel. Subjects could listen to each token maximum 2 times.

## 3. Results

### 3.1. Prompt sentences

In terms of quality, vowels in the target syllables of the prompt sentences display expected variation with long and stressed vowels being more peripheral, especially in F2. Figure 1 shows the vowels in F1-F2 space of perceptually natural bark scale. Predictable variability due to coarticulation can be seen.

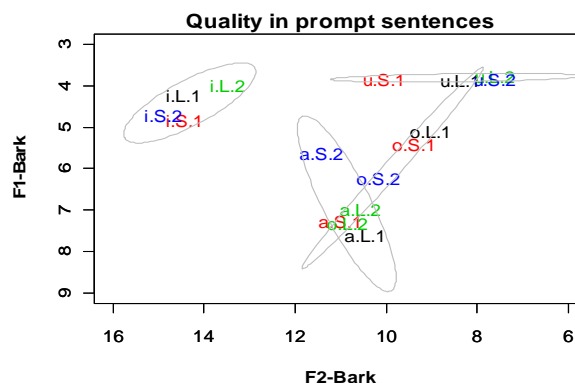


Figure 1: F1 and F2 at the midpoint of the four target vowels (L-long, S-short, 1-stressed, 2-unstressed).

In terms of duration, all short vowels are well separated from the long ones; the ratios are at or below 1.5. Minima were comparable to data extracted for intrinsic durations while maxima were significantly lower, all at or below 100ms.

### 3.2. Perception of vowel quantity

The results are illustrated with binomial regression curves in Figure 2. Several patterns can be observed. First, quantity contrast is robustly present in the perception of Slovak vowels since the steep slopes of the curves are present in all panels. Comparing this with Hungarian, stressed and unstressed /a/ in [8] did not show an S-shape within the range of durations used in that experiment. Hence, Hungarian subjects did not use duration but quality in differentiating the long/short /a/ counterparts.

Second, comparing the stressed (top) and unstressed (bottom) vowels, we see that the flat portion of the curve in the upper-left part of the panel (perception of short vowel) is much shorter than the flat portion in the bottom right corner for unstressed vowels than for stressed ones.

This applies to all four vowels and can also be seen by comparing the inflection points. It means that subjects need shorter duration to perceive vowels as long in unstressed

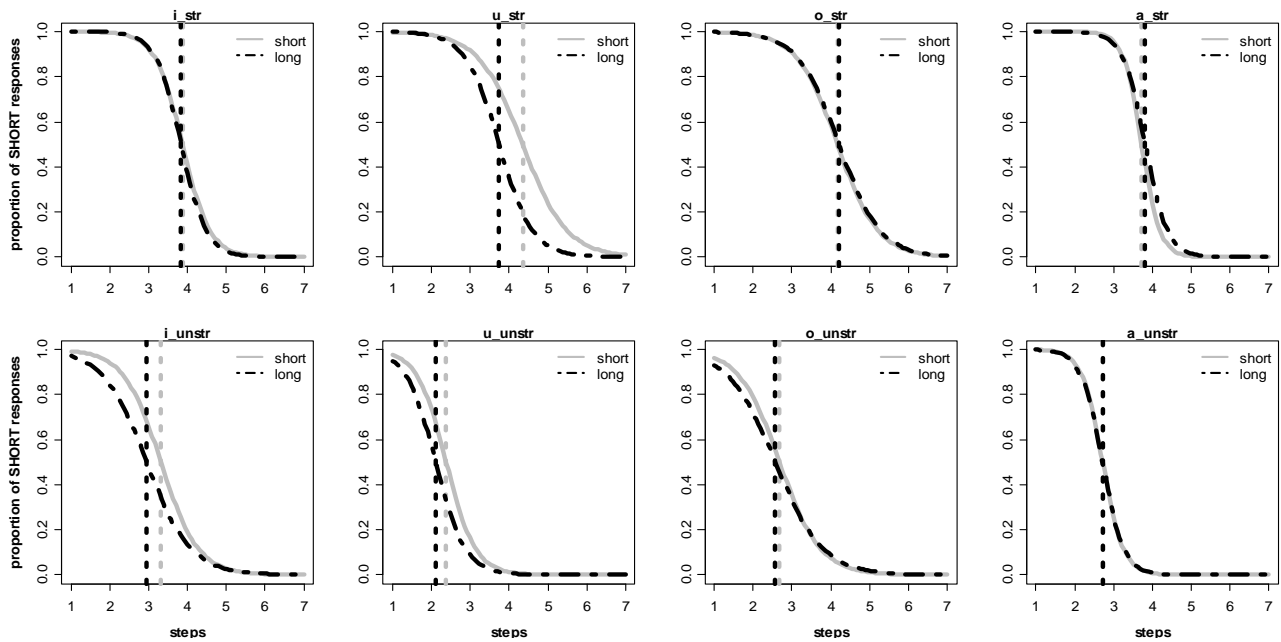


Figure 2: Mean binomial regression curves based on 17 subject responses plotted as the proportion of identifying short vowels depending on the quantity continuum (1-shortest, 7-longest) and divided by vowel type, stress, and phonemic length. Vertical dashed lines indicate inflection points marking chance identification between two quantity categories.

syllables than in the stressed ones. Hence, subjects expect shortening in unstressed syllables due to prosodically motivated undershoot and take it into account in their identification of phonological quantity. Testing this observation quantitatively, a repeated-measures MANOVA was performed on all data with stress, vowel type, and quantity as independent variables, inflection point as a dependent variable, and participant as a within-subject factor. The effect of stress was highly significant ( $F = 234.4$ ,  $p < 0.001$ ), but it also interacted with vowel type significantly ( $F = 29$ ,  $p < 0.001$ ). Separate RM MANOVA for each vowel showed that the greatest effect of stress was on /u/, then /a/, then /o/, and /i/, all of which were significant at  $p < 0.001$ .

Third, the quality difference between the long and short counterparts of each vowel plays a small role in quantity distinction. This can be seen by comparing the inflection points that indicate 50% (chance) identification between two quantity categories in each plot. The clearest difference is in stressed /u/ where the originally long vowel is perceived as long at shorter vowel duration than the same vowel created on the basis of the originally short vowel. A similar pattern can be observed for unstressed /i/ and /u/ vowels. In the same RM MANOVA test, the main effect of quantity, its interaction with vowel type, as well as three-way interaction with stress were significant ( $F = 30.2$ ,  $p < 0.001$ ,  $F = 31.8$ ,  $p < 0.001$ , and  $F = 10.5$ ,  $p < 0.001$  respectively). In separate tests for the four vowel types, the main effect of vowel quantity for /u/ and /i/ were observed ( $F = 103$ ,  $p < 0.001$  and  $F = 18.7$ ,  $p < 0.001$  respectively), but for both vowels also quantity~stress interactions were significant ( $F = 26.7$ ,  $p < 0.001$  for /u/,  $F = 6.9$ ,  $p = 0.018$  for /i/). Hence, the quality of the original vowels /u/ and /i/ affects perception in the expected direction: the originally long vowel is perceived as long at shorter duration than the same vowel created on the basis of the originally short vowel. However, this effect is stronger in the

stressed than unstressed /u/ and applies only to unstressed /i/.

Fourth, the surprising finding is the contrast between /i/ and /u/ in the direction of the interaction between stress and quantity. While for /u/ the difference between the inflection points of the long and short vowels is greater in the stressed vowels than the unstressed ones, the opposite pattern is observed for /i/; both effects are significant as shown in the tests reported above. This reflects the quality differences in the original data in Figure 1: larger for stressed /u/ and for unstressed /i/ than unstressed /u/ and stressed /i/.

Finally, we observe that the slopes of the curves and inflection points for vowel types are relatively similar. This comparison, however, should be done with caution since the four vowels did not have the same duration of the steps in creating the continua. It seems that the steepness differences follow the differences in steps: /a/ has the steepest curves, then /i/ and then /o/ and /u/. This might suggest that our effort to control for the intrinsic vowel duration by using the range of quantity variation from a production corpus might have made the difference between the vowel ranges too big. On the other hand, our result might also signal that perceived robustness of the categorical boundary for /a/ is the greatest, then for /i/, and the other two vowels. This conclusion is supported also by observed frequencies of Slovak phonemes reported in [2] in which the long /a/ is the most frequent at 2.07, long /i/ is second with 1.99, the third long /u/ with a significant lag at 0.9, and long /o/ with a minimal frequency of 0.01. A mixed models test [10] allows filtering out the effects of more than one categorical variable. In this test with the slope of S-curves as a dependent variable, quantity and stress as fixed factors, and subject and vowel type as random factors, stressed vowels produced steeper slopes than unstressed ones in the pooled data ( $F = 8.7$ ,  $p \sim 0.01$ ). In tests for separate vowels, this effect was robustly present for /i/, less so for /a/, and not significant for /o/. However, the opposite direction of the effect was observed for /u/: slopes for the stressed vowels

were less steep than for the unstressed vowels. This result suggests a decrease in the identification of unstressed /i/ and /a/ due to their prosodic weakness, and the *increase* in the identification of unstressed /u/.

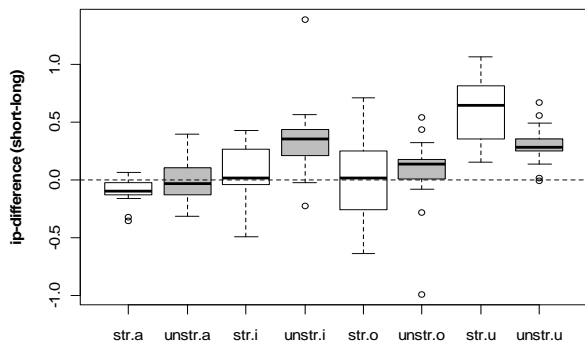


Figure 3: Differences between the inflection points from short and long vowels for all subjects divided by vowel type and stress.

Figure 3 is a slightly different way of looking at the data that allows examining between-subjects variability in the effects of quality on the perception of vowel quantity. The greater the difference from zero, the greater effect of quality on quantity perception. We see that for /a/, /o/, and stressed /i/, the difference is about zero. Recall that stressed /a/, /o/, /i/ had very similar F1 and F2 frequencies, and thus, the absence of the effect of quality on their quantity identification might be due to stimuli. However, this applies also to unstressed /a/ and /o/, for which the long and short counterparts in the stimuli were separated much better. On the other hand, unstressed /u/ has virtually identical qualities for long and short counterparts, but the difference between the inflection points suggests that subjects did not rely entirely on the duration cue.

Figure 3 also shows that for three vowels except /a/, the variance in the unstressed vowels is smaller than in the stressed ones; the opposite applies to /a/. A Levene test for homogeneity of variances shows that greater variance in stressed than unstressed vowels reaches only a tendency in the pooled data,  $F = 3.5$ ,  $p = 0.06$ , which is due only to the significant variance difference between stressed and unstressed /u/;  $F = 4.3$ ,  $p = 0.05$ . This suggests that subjects were more consistent in perceiving the boundary between long and short /u/ in the unstressed than in the stressed syllables.

#### 4. Discussion and conclusion

H&H theory [4] predicted steeper S-curve slopes for stressed vowels while the functional load characteristics predicted steeper curves for unstressed vowels. Our data showed that Slovak vowels do not behave uniformly with respect to this issue. Both /i/ and /a/ support the H&H prediction, /o/ does not prefer either, and /u/ tends to support the functional load prediction. This is because /u/ shows smaller variance and steeper slopes in the identification of unstressed than stressed vowels. Moreover, /u/ is the only vowel that shows the effect of quality on the perception of its quantity in both stressed and unstressed vowels. These findings might also be related to weak jaw support for this vowel in general as well as for the quantity distinction observed in [5].

Despite this, Slovak shows minimal deterioration of categorical difference between stressed and unstressed vowels,

notwithstanding clear shortening in unstressed syllables. Therefore, it seems that the phonetic requirement for weaker unstressed vowels and greater functional load placed on these unstressed vowels tend to cancel each other out: weaker functional load is boosted by the phonetic salience of the first syllables while heavier load in unstressed syllables is countered by phonetic neutralization of the contrast.

Further exploration is needed for the effect of coarticulation in vowel quantity for the perception of phonemic quantity. Some, but not all, effects in this perception study can be linked to the coarticulatory patterns in the stimuli (e.g. the difference between short and long stressed /u/ in their coarticulation to the following /i/, or between two unstressed /i/ in their coarticulation to the preceding vowel).

Languages surrounding Slovak (mostly Hungarian, but recently also Czech) seem to be losing quantity distinctions especially in high vowels, and, in some vowel classes, quantity distinction are becoming substituted by quality distinction. Given this, we wanted to test for a similar trend in Slovak. All our subjects were younger than 25 and if a change in this direction is taking place, we should notice it in this age group. Our data suggest that the quantity-quality system especially in high vowels may become imbalanced.

#### 5. Acknowledgements

This publication is the result of the project implementation: Technology research for the management of business processes in heterogeneous distributed systems in real time with the support of multimodal communication, ITMS 26240220064 supported by the Research & Development Operational Programme funded by the ERDF, was supported in part by the Alexander von Humboldt grants to both authors and the German Research Council and the Hungarian Scientific Research Fund (OTKA) to the 2<sup>nd</sup> author. Thanks also to Ch. Draxler, IPS Munich, for the data-collection tool.

#### 6. References

- [1] Beňuš, Š., Mády, K., "Effects of lexical stress and speech rate on the quantity and quality of Slovak vowels", Proc. of 5<sup>th</sup> Speech Prosody, 1-4, 2010.
- [2] Kráľ, A. and Sabol, J., "Fonetika a fonológia", Bratislava: Slovenské pedagogické nakladateľstvo, 1989.
- [3] Hockett, Ch. F., "A Manual of Phonology", Indiana University Publications in Anthropology and Linguistics 11, 1955.
- [4] Lindblom, B., "Explaining phonetic variation: a sketch of the H and H theory", in Hardcastle, W. J., Marchal, A. Speech Production and Speech Modelling, Kluwer, pp. 403--439, 1990.
- [5] Beňuš, Š., Pouplier, M., "Jaw movement in vowels and liquids forming the syllable nucleus", Proc. of 12th Interspeech, 389-192, 2011.
- [6] Beňuš, Š., Cerňák, M., Darjaa, S., Rusko, M., Trnka, M., "Adapting Slovak ASR for native Germans speaking Slovak. Proc. of EMNLP, 2011.
- [7] Mády K. and Reichel, U.D., "Quantity distinction in the Hungarian vowel system—just theory or also reality?", Proc. 16th ICPHS, 1053–1056, 2007.
- [8] Mády, K., "Shortening of long high vowels in Hungarian: a perceptual loss?", Proc. Sociophonetics at the crossroads of speech variation, processing and communication, 2010.
- [9] Podlipský, V. J., Skarmitzl, R., and Volín, J., "High Front Vowels in Czech: a Contrast in Quantity or Quality, Proceedings of Interspeech, 132-135, 2009.
- [10] Baayen, R. H. Analyzing Linguistic Data. A Practical Introduction to Statistics Using R Cambridge: CUP, 2008.