

High Front Vowels in Czech: a Contrast in Quantity or Quality?

Václav Jonáš Podlipný¹, Radek Skarnitzl², Jan Volín²

¹ Department of English and American Studies, Philosophical Faculty, Palacký University Olomouc

² Institute of Phonetics, Faculty of Arts, Charles University in Prague

podlip@ffnw.upol.cz, radek.skarnitzl@ff.cuni.cz, jan.volín@ff.cuni.cz

Abstract

We investigate the perception and production of Czech /ɪ/ and /i:/, a contrast traditionally described as quantitative. First, we show that the spectral difference between the vowels is for many Czechs as strong a cue as (or even stronger than) duration. Second, we test the hypothesis that this shift towards vowel quality as a perceptual cue for this contrast resulted in weakening of the durational differentiation in production. Our measurements confirm this: members of the /ɪ-/i:/ pair differed in duration much less than those of other short-long pairs. We interpret these findings in terms of Lindblom's *H&H theory*.

Index Terms: vowel quality, vowel quantity, duration, Czech

1. Introduction

Several studies suggest that in different languages classified as quantity languages, the role of segmental duration may vary depending on whether short and long categories are distinguished by duration alone or not. For instance, Kirmse *et al.* [1] recently showed an enhancement of phonetic processing of vowel duration (measured by the mismatch negativity) in Finns as opposed to Germans which they ascribed to the lower relevance of duration as a distinctive cue in German. Mády and Reichel [2] reported that in Hungarian, vowel quantity and vowel quality interact closely in perception; the authors concluded that quantity in Hungarian is not as important as it had been considered. Pind [3] found that in Icelandic, for the [ɛ]-[ɛ:] vowel pair differing in quality (despite the transcription), this quality difference can work as a stronger perceptual cue than duration.

The Czech language has five vowel pairs distinguished by phonological length. Traditionally, the long vowels had been described to be approximately twice as long as the short vowels [4]. However, our informal observations suggested that this may not be the case in the /ɪ-/i:/ pair. This is perhaps because there is also a difference in quality between its members. While this qualitative difference, and therefore an exceptional status of /ɪ-/i:/ among Czech short-long vowel pairs, has been recognized in literature [5: 33] (even though quite reluctantly by prescriptive linguists) its potential contribution to the contrast has not been meaningfully considered.

Recall the finding by Pind about Icelandic cited above. In a subsequent paper [6], Pind tested the hypothesis, based on Lindblom's *H&H theory* [7], that since Icelandic [ɛ] and [ɛ:] have different spectra, a weakening of the temporal differentiation between them may occur in production. This was however not confirmed in Pind's production experiment. In the present study, we set out with the hypothesis that in the case of Czech /ɪ/ and /i:/, the (so far empirically not confirmed) usage of spectral cues for perception may have indeed resulted

in a diminished long/short vowel ratio in production when compared to other Czech short-long vowel pairs where reliable spectral cues are absent.

In order to test this hypothesis, perception of the Czech /ɪ-/i:/ contrast had to be examined first (Experiment 1). Subsequently, we measured the duration of these two vowels as well as other Czech short and long vowels in natural speech production, so as to compare the ratios of long/short vowel durations (Experiment 2).

2. Experiment 1: perception

As mentioned above, both temporal and spectral differences exist between the Czech 'long' and more close /i:/ and the 'short' and more centralized /ɪ/ [5]. It has not yet been empirically tested whether or not the spectral difference is employed by Czech listeners as a perceptual cue. The purpose of this experiment was thus to assess the relative weight of duration and spectrum as cues to the /ɪ-/i:/ contrast in native Czech listeners.

2.1. Method

2.1.1. Stimuli

The stimuli were a set of 36 syllables of the shape [bVt] where the vowel ranged from [ɪ] to [i:]. The endpoints thus formed the minimal pair *byt* [bɪt] 'an apartment' vs. *byť* [bit] 'to be'. As schematized in Figure 1, the stimuli produced an array of 6 durations (94.6 – 179 ms, step of 16.87 ms) by 6 spectra (endpoints 175-357-2120-2680 and 175-207-2450-3200 [F0-F1-F2-F3 in Hz]; intermediate formant values were interpolated in Mel). These endpoints, corresponding to typical Czech male [ɪ] and [i:] in that context, were chosen rather than values spanning an equal number of just-noticeable-difference units (JNDs) in both dimensions: the duration span was 11.8 JNDs and the spectrum span was only 6.1 JNDs (see [8] for the conversion procedure used). Therefore, if listeners were observed to categorize the spectrally-differing stimuli differently, despite the smaller JND span, it would reflect a robust reliance on spectrum as a perceptual cue.

All stimuli were created by resynthesizing a single sentence *Řekni byt znovu* 'Say apartment again' pronounced by a Czech male speaker (the carrier sentence was kept in the stimuli). Both the resynthesis, which preserved the formant contours of the original token, and the duration manipulation was performed in Praat [9] (see Praat manual entry 'Source-filter synthesis 4. Using existing sounds' for a description of the resynthesis procedure).

The stimuli (ordered from open to close qualities and from short to long durations) are given in the supplementary sound file (.wav).

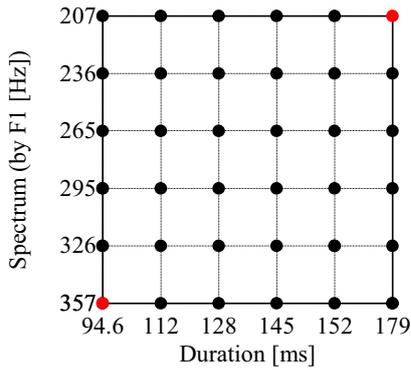


Figure 1: Stimulus array covering the spectral (F1, F2, F3) and durational span between [ɪ] (lower left corner) and [i:] (upper right corner).

2.1.2. Procedure

We conducted a categorization task using a 2-alternative-forced-choice paradigm: participants heard a stimulus (via circumaural headphones) and with a computer mouse they clicked on one of two response buttons marked *byt* and *být* (listeners could replay a stimulus twice at most). After the response, another stimulus was presented (ISI 1.1 s). Each of the 36 stimuli was presented 6 times in pseudorandom order (repetition only after all stimuli had been presented in random order). There were thus 216 trials divided into 3 blocks of 72 trials (listeners watched a short Mr. Bean video during the pauses).

2.1.3. Participants

Fifty-one native Czech speakers, 41 female and 10 male, aged 19–25 (mean 20.1, standard deviation 1.4), took part in the experiment. They were students at Palacký University and were compensated for participation. They all grew up in a Czech monolingual environment. None of them reported any hearing or language-related impairment.

It was hypothesized that regional differences may exist in the perception of the /ɪ-/i:/ contrast. While our participants were from various parts of the Czech Republic, we sorted them in two gross regional groups: Bohemians ($n = 20$) and Moravians ($n = 31$). The reason for this distinction is that Moravia, the eastern part of the Czech lands, generally preserves older forms of pronunciation; Bohemians live in the western part of the country and represent more progressive forms of pronunciation.

2.2. Results

Figure 2 shows the mean proportions of *být* /i:/ and *byt* /ɪ/ responses pooled over all participants. It is apparent that on average, spectrum was as much relied on by the listeners as duration, and that the category boundary between /ɪ/ and /i:/ (i.e. the 50 % labelling contour line) cut diagonally through the duration-by-spectrum stimulus space.

It must be stressed, however, that Figure 2 conceals a great deal of variation between listeners. To quantify the slope of the category boundary in the duration-by-spectrum space for each listener individually, two methods were used.

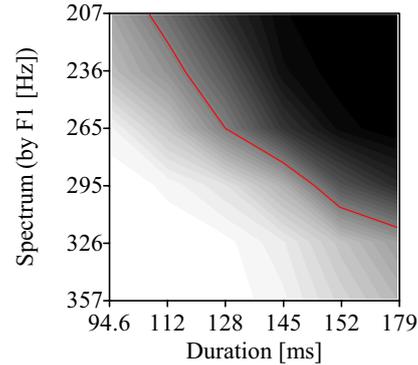


Figure 2: Mean proportions of *být* /i:/ (dark) and *byt* /ɪ/ (light) responses (pool over all participants). The 50 % contour line is highlighted. Dimensions correspond to the 6×6 stimulus array shown in Figure 1.

First, we computed the ratio between the reliance on duration (defined as the difference between the proportions of /i:/-responses to the longest and to the shortest stimulus) and the reliance on spectrum (defined analogically); this metric was proposed by Boersma and Escudero [10]. The results, broken down by listeners' region of origin, are shown in Figure 3. One-way ANOVA revealed a significant main effect of region ($F[1, 49]=12.45, p<.001$): in Boersma and Escudero's terms [10], Moravians relied on 'duration and spectrum' while Bohemians relied on 'spectrum and duration' on average.

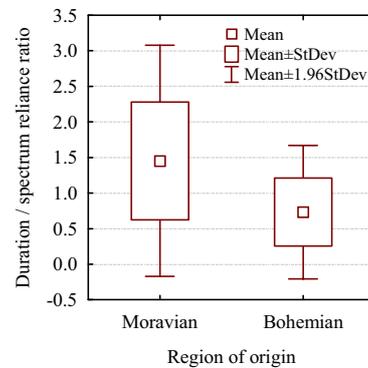


Figure 3: Boxplots of the duration / spectrum reliance ratios [10] of the 51 listeners categorizing Czech /ɪ-/i:/. Distributions are broken down by listeners' region of origin.

Second, since the duration/spectrum reliance ratio is only based on the data points at the edges of the stimulus space, we fitted logistic regression models to each listener's proportion of /ɪ/ responses, using Morrison's software [11], as a supplementary analysis. The model, shown in Equation 1, included the intercept α and the coefficients β_{dur} and β_{spec} tuned by the durational and spectral properties of the stimuli (x_{dur} and x_{spec}) respectively.

$$\ln \left(\frac{p(\text{/ɪ/} | x_{dur}, x_{spec})}{1 - p(\text{/ɪ/} | x_{dur}, x_{spec})} \right) = \alpha + \beta_{dur} x_{dur} + \beta_{spec} x_{spec} \quad (1)$$

We also compared this model with models which excluded β_{dur} or β_{spec} in a ΔG^2 test. For all 51 listeners, both duration and spectrum significantly improved the goodness of fit of the logistic regression model ($p < .0001$). This corroborates the conclusion of the reliance-ratio analysis that listeners relied both on duration and spectrum when categorizing Czech /ɪ/ and /i:/. Figure 4 plots the β_{dur} values against the β_{spec} values of individual listeners. Again, it is obvious that there was a good deal of between-subject variation. However, the Bohemians tended to have higher β_{spec} values and β_{dur} values closer to zero than the Moravians.

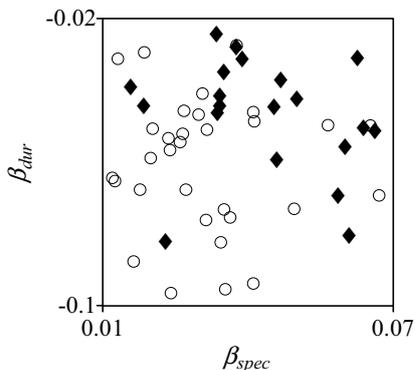


Figure 4: A scatterplot of the β_{dur} and β_{spec} outputs of the logistic regression model fitted to each listener's proportion of /ɪ/ responses. Bohemians: black diamonds, Moravians: white circles.

2.3. Discussion

In sum, it seems that the qualitative difference between Czech /ɪ/ and /i:/ is utilized by listeners as a perceptual cue. Having said that, we have to add that not all listeners relied to an equal degree on duration and on spectrum when identifying the two vowels. Roughly speaking, listeners from Bohemia were more likely to favor spectrum over duration than listeners from Moravia who showed the opposite tendency.

It appears that for most Czechs, the historically durational contrast has been redefined as a contrast integrating both qualitative and quantitative differences. Such cue integration is very common in speech perception (e.g. [12]). One way of accounting for this redefinition is to say that a listener-driven sound change [13] has been taking place in Czech. Although we do not have any direct evidence to support this, we speculate that the pattern observed in Moravians represents an older stage in the development of the high front vowel contrast. The /ɪ/-/i:/ contrast would not be the only case when Bohemian pronunciation seems to be more progressive than Moravian pronunciation.

3. Experiment 2: production

The aim of the second experiment was to determine whether the redefinition of the /i:/ vs. /ɪ/ contrast, proposed on the basis of the results of Experiment 1, is reflected in durational differences found in natural production data. As mentioned in the Introduction, phonologically long vowels in Czech have been typically described as being, on average, twice as long as phonologically short vowels [4], with no exception indicated for the /ɪ/-/i:/ pair.

3.1. Method

The speech material was drawn from news bulletins read by six professional newscasters of the public broadcaster Czech Radio (three females, three males). The recordings, each approximately four minutes long, were manually segmented by the second and third author of this study. Uniform, phonetically motivated segmentation criteria were applied. The segmentation was carried out for general purposes of corpus-building, unrelated to this study. The duration of all vowels (over 7,100 items in total) was measured in Praat [9].

3.2. Results and discussion

Table 1 shows the mean durations and their standard deviations of all ten Czech monophthongs and the diphthong /ou/. Note that the long /o:/ is quite rare in Czech, as it only appears in loan words. That is why duration data for the diphthong /ou/ are also listed in Table 1, albeit in grey. The most relevant information for the present study can be found in the last column of Table 1. It is the duration ratio of each of the long-short pairs.

First of all, it is obvious that none of the long vowels seems to be twice as long as the respective short vowel, with the exception of the /ou/-/o/ pair. What is more interesting, however, is the marked difference between the long/short vowel ratio of the front close vowels and all the other vowel pairs: the phonologically long /i:/ is only about 30 % longer than the short /ɪ/, while the ratios for the other vowel pairs are 1.6 or higher.

Table 1. Mean durations and long/short vowel duration ratios.

vowel	n	duration [ms]	SD [ms]	ratio V: / V
ɪ	1,049	53.5	19.9	1.29
i:	691	68.9	36.3	
ɛ	1,595	53.5	16.2	1.72
ɛ:	211	91.8	37.1	
a	1,217	63.1	20.9	1.79
a:	436	113.0	26.3	
o	1,259	51.7	15.4	1.73
o:	5	89.3	17.7	
ou	96	102.8	26.4	1.99
u	423	57.3	22.9	1.60
u:	133	91.4	49.5	

As it is well known that phrase-final vowels undergo final lengthening, we examined separately vowels in phrase-final syllables and vowels in all other positions. Figure 5 shows the durations and the long/short ratios for non-final and final vowels. First, we can see that vowels in phrase-final syllables are longer than vowels within a phrase, but the individual vocalic pairs do not behave identically. More importantly, however, the ratios between the durations of long and short vowels are considerably different. When we compare these ratios (i.e. ratios broken down according to the position in the phrase) with the overall results presented in Table 1, we can see the greatest change for the /ɛ/-/ɛ:/ pair. This seems to be caused by the occurrence of a great portion of phrase-final /ɛ:/-items in feminine adjectives, which have no syllabic coda in

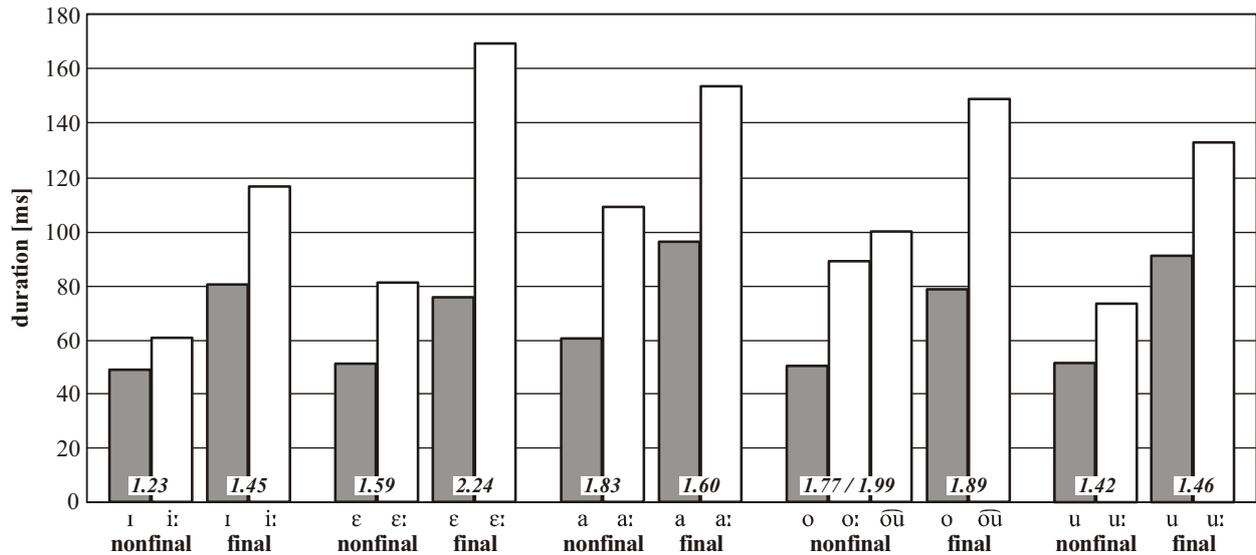


Figure 5: Mean durations and long-short vowel duration ratios for vowels in phrase-final and non-final syllables.

Czech. Indeed, vowels which appear in adjectival endings, /i: ε: a:/, seem to deviate most from the averaged values given in Table 1. The most important finding with respect to the /i/-/i:/ pair is the fact that /i:/ is extremely short in non-final syllables, while the contrast between /i/ and /i:/ is more marked in final syllables.

4. General discussion

We suggest that the quantitative symmetry of the Czech vowel system has been eroded somewhat. This is because, on the one hand, listeners employ both temporal and spectral cues for distinguishing the members of the /i/-/i:/ contrast, and on the other hand, the comparison of the ratios of long and short vowels showed that /i:/ is on average only 30 % longer than /i/. The phonologically long /i:/ can “afford” to be comparatively shorter precisely because, due to the difference in spectral profile, there is little risk of confusion with /i/.

Unlike Pind’s [6] findings about Icelandic [ε] and [ε:], our results seem to be interpretable in terms of Lindblom’s H&H theory [7]. Within this framework, the conclusion of the previous paragraph could be rephrased in the following way: due to the fact that Czech /i/ and /i:/ differ in quality and that this difference is utilized by listeners as a perceptual cue, Czech speakers can “afford” to shorten the phonologically long vowel /i:/. This economy on the part of the speakers will not endanger correct perception.

5. Acknowledgements

We thank Geoff Morrison for comments on the perception experiment.

The second and third author gratefully acknowledge support of the European grant MRTN-CT-2006-035561 “Sound to Sense” and of the grant VZ MSM 0021 620825.

6. References

- [1] Kirmse, U. et al., “Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience”, *Int J Psychophysiol*, 67(2): 131-143, 2008.
- [2] Mády, K. and Reichel, U. D., “Quantity distinction in the Hungarian vowel system – just a theory or also reality?”, *Proc 16th ICPHS*, 1053-1056, Saarbrücken, 2007.
- [3] Pind, J., “Spectral factors in the perception of vowel quantity in Icelandic”, *Scand J Psychol*, 37(2):121-131, 1996.
- [4] Palková, Z., *Fonetika a fonologie češtiny*, Praha: Karolinum, 1994.
- [5] Petr, J. et al., *Mluvnice češtiny I*, Praha: Academia, 1986.
- [6] Pind, J., “Speech segment durations and quantity in Icelandic”, *J Acoust Soc Am*, 106(2):1045-1053, 1999.
- [7] Lindblom, B., “Explaining phonetic variation: a sketch of the H&H theory”, in W. J. Hardcastle, A. Marchal [Eds], *Speech production and speech modeling*, 403-439, Dordrecht: Kluwer, 1990.
- [8] Morrison, G. S., “L1-Spanish speakers’ acquisition of the English /i/-/I/ contrast II: Perception of vowel inherent spectral change”, *Lang Speech*, 52(4): in press.
- [9] Boersma, P. and Weenink, D., “Praat: doing phonetics by computer (Version 4.6) [Computer program]”, retrieved May, 2007, <http://www.praat.org/>.
- [10] Boersma, P. and Escudero, P., “Measuring relative cue weighting: A reply to Morrison”, *Stud Second Lang Acq*, 27(4):607-617, 2005.
- [11] Morrison, G. S., “Logistic regression software for speech perception data [Computer program]”, retrieved March 7, 2009, <http://geoff-morrison.net/>.
- [12] Repp, B. H., “Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception”, *Psychol Bull*, 92(1):81-110, 1982.
- [13] Ohala, J. J., “The listener as a source of sound change”, in C.S. Masek, R. A. Hendrick and M. F. Miller [Eds], *Papers from the Parasession on Language and Behavior*, 178-203, Chicago: Chicago Ling. Soc., 1981.