euronoise

Vowel duration as a cue for coda voicing and the perception of second-language vowel quantity

Vaclav Jonas Podlipsky

Dept. of English and American Studies, Palacky Univ., Krizkovskeho 10, 77180 Olomouc, Czech Republic
vaclavjonaspodlipsky@centrum.cz

When acquiring the phonology of a second language (L2), a learner may have to dissociate a perceptual cue from what it marked in the first language (L1) and attach it to another linguistic entity. This study examined the acquisition of Czech vowel quantity by native speakers of American English. In Czech, vowel duration is reserved for cuing short/long vowel (V/V:) contrasts, while in English it is an important cue for coda obstruent voicing. It is not certain whether in Czech, as in many other languages, voiced codas lengthen the preceding vowels; if they do, a vowel should be more likely to be perceived as short before a voiced and as long before a voiceless obstruent. A perceptual V/V: categorization experiment showed that for Czech listeners (n=54) the V/V: boundary came slightly (3.93 ms) but consistently ($p<.001$) later before a voiced coda. If English learners transfer L1 perceptual strategies to their L2 Czech, the V/V: boundary shift should be even larger. However, this study found no significant effect of coda voicing on the perception of vowel quantity for the non-native listeners. It is concluded that the learners redefined the value of vowel duration as a perceptual cue.

1 Introduction

Accurate perception of second language (L2) speech implies accurate discrimination of L2 speech sound categories. In order for a learner to accomplish accurate discrimination, acoustic cues forming the basis of contrast between categories need to be discovered. The number of different acoustic parameters that can be used to contrast linguistic units is constrained by the physiology of speech production and perception. Therefore a new 'attunement' [1, 2] of cues used both in the first language (L1) and in L2, but for different purposes in each, may be a necessity for successful L2 acquisition (the premise being that bilinguals have a common L1 and L2 phonological space [3]).

One acoustic dimension that can serve as a cue for perceptual categorization of speech sounds is their duration. Speech is a series of events that happen in time and duration is therefore naturally inherent in all languages. However, its specific role is not cross-linguistically constant. A number of studies (e.g. [2, 4]) have examined if accurate perceptual distinction of non-native vowel quantity contrasts can be achieved by learners in whose L1 vowel duration is not used contrastively at the segmental level but it is utilized suprasegmentally. Some studies specifically addressed the question of how L2 vowel quantity interacts with linguistic dimensions (such as stress [5] or accent [6]) to which learners may attach vowel duration because of their L1s. In these studies, a tendency was observed in learners to transfer their L1-like perceptual strategies into L2.

Apart from marking stress for instance, vowel duration can serve as a cue for the perception of the voicing value of the following obstruent. The tendency for a vowel to be shorter when it is in a syllable closed by a voiceless (as opposed to voiced) consonant, being probably a universal articulatory by-product [7], seems to have acquired different importance within the phonologies of different languages. In English for example, this effect is relatively large and several authors (e.g. [8]) suggested that it has been phonologized to enhance the voicing contrast of the obstruent in the coda, possibly even outweighing the phonetic voicing of the consonant itself [9]. On the other hand, in quantity languages where vowel duration is reserved for contrasting vowel length, it is not readily used allophonically or suprasegmentally [10] (unless the 'short/long' vowel opposition is based partially on qualitative differences like in German or Dutch where phonologically voiced codas have indeed been shown to lengthen preceding vowels [11]).

The goal of the present study was to test how English learners of Czech perceive non-native vowel quantity. More specifically, the aim was to determine if learners would transfer L1 perceptual strategies to the L2 and if their perception of L2 short/long vowel contrasts would be thus affected by voicing of the following consonant.

It has to be noted that it is not clear whether in Czech any interaction between vowel quantity and coda voicing exists. In fact, few studies have explored if coda voicing exerts any effect on nuclear vowel duration in Czech at all. Machač and Skarnitzl [12] analyzed naturally-produced vowel-consonant (VC) sequences and showed that only some vowels were significantly longer before (phonetically) voiced consonants, and that only applied to some places of articulation. In addition, heterosyllabic VC sequences were not treated separately from tautosyllabic sequences in their study. Their findings do not therefore allow drawing reliable conclusions. However, what is more interesting, if systematic variation in vowel duration caused by coda voicing really exists in Czech, it is likely to interact with vowel quantity. It is clear that quantity contrasts are defined in relative rather than absolute terms; in Japanese the short/long category boundary shifts with differing speaking rate [13]; in Czech vowels of ambiguous duration are more likely to be perceived as short in open syllables and as long in closed syllables [14].

It is evident now, that to be able to decide if English learners of Czech transfer the usage of vowel duration as a perceptual cue for coda voicing to their L2 Czech and adjust their perception of Czech vowel length accordingly, it was first necessary to obtain baseline data about native Czech listeners. In English, coda voicing causes quite a large vowel-duration difference (10-20 ms in non-phrase-final syllables [8]). For this reason, it was hypothesized that transfer would result in the non-native listeners showing a larger short/long vowel boundary shift than any potential shift observed in the native speakers.

2 Method

Two tests were created to compare the perception of a Czech short/long vowel contrast located before a tautosyllabic voiceless coda with the perception of this contrast before a tautosyllabic (phonetically) voiced coda.

2.1 Stimuli

Two separate vowel-duration continua (each having 8 members) were used in this study, one continuum involving

a voiceless-coda context and the other a voiced-coda context.

In the first continuum, the initial stimulus was the nonsense word [tapka] produced naturally in the carrier phrase *Slyším ... jasně* ('I can hear … clearly') by a male Czech speaker. The vowel [a] located in the first syllable was the target segment and it lasted 64 ms. In the subsequent stimuli in the continuum, this vowel was lengthened (by multiplying fundamental cycles in its steady state) in 7 steps with a constant increment of approximately 12.06 ms (two cycles) so that the final 8th stimulus was the (nonsense) word [ta:pka] with the long vowel lasting 148.4 ms. The durations of the target vowel in the intervening stimuli were 76.1, 88.1, 100.2, 112.3, 124.3, 136.4 ms respectively.

The second continuum differed from the first one in the voicing of the stop following the manipulated vowel ([tapka] – [ta:pka] vs. [tabga] – [ta:bga]). A disyllabic word was necessary because Czech has final devoicing and a voiced obstruent in the coda position is voiced phonetically only if the following segment is voiced. Importantly, there was no difference whatsoever between the two continua in the target vowel durations. This was accomplished by creating the voiceless-coda continuum first and then splicing the CV(:) of each stimulus (i.e. the two initial segments) and the sequence [bga] excised from the naturally-produced nonsense word [tabga]. Having identical vowel durations in the two continua allowed making conclusions about the effect of the difference in the following consonant.

The manipulated word was selected to ensure that both endpoints of both continua were nonsense words rather than real words and that confounding lexical effects were thus ruled out. Also, the nonsense words were chosen to avoid ambisyllabicity of the consonant following the manipulated vowel – although Czech phonotactics are considerably permissive (e.g. [15]), no Czech word begins with the cluster #*pk*- or #*bg*-.

The stimuli were presented along with the original carrier phrase to achieve a better naturalness of the listening material and to equalize the speech tempo. (The stimuli can be found in the electronic version of the proceedings as audio files *tapka-taapka.mp3* and *tabga-taabga.mp3*.)

2.2 Procedure

Two 2-alternative forced-choice identification tests were conducted, one for each continuum. Each stimulus was presented 10 times and there were thus 80 trials within each test, the order of which was randomized and different for every listener.

Subjects listened to the stimuli via headphones in a quiet room and responded by clicking on one of the two buttons displayed on the screen of a laptop computer using a mouse. The buttons were marked *tapka* and *tápka* in the first test, and in the second test *tabga* and *tábga* (the stroke above a vowel symbol indicates vowel length in standard Czech orthography).

Before each test, every listener completed a training task at least once with a subset of the stimuli (the endpoints were included). Each person completed both tests, with at least a 20-minute break between the tests, but the order of the tests was counterbalanced within each group of subjects.

2.3 Subjects

There were two groups of participants – 54 native Czech speakers and 16 American English learners of Czech. None of the subjects reported any hearing problems or language dysfunctions.

Fifty-four L1 listeners, aged between 20 and 46 years and all coming from the Olomouc region of the Czech Republic, took part in the experiment. Preference was given to monolinguals and none of the participants reported advanced command of a foreign language.

	Mean	SD
Months of learning	14.5	6.7
Months of residence	12.3	6.7
Age of onset of learning	19.6	1.4
% of using L2 (vs. L1)	44.4	12

Table 1 Information about the non-native speakers of Czech

Information about the sixteen L2 listeners is summarized in Table 1. They were all native speakers of American English, aged between 19 and 24, who had come to the Czech Republic as religious missionaries. As a group, they seemed to have a strong motivation to learn Czech, judging by the relatively high self-estimated percentage of using Czech (as opposed to English) and by their good ability to hold a conversation in Czech (at least half of the interview with these subjects was carried out in the L2).

3 Results

The number of subjects in the native and the non-native group was not balanced. This did not present a problem however because only within-group statistics were necessary to address the experimental question.

3.1 Baseline Czech results

Identification curves averaged across the native Czech listeners are shown in Fig. 1. When both curves of each subject were submitted to repeated-measures ANOVA, a highly significant (p<.00001) interaction between vowel duration and test condition (i.e. coda voice value) was revealed. Post-hoc tests compared mean responses to individual stimuli and showed that responses to the 4th stimulus (quantitatively the most ambiguous stimulus) differed significantly (p<.001 Scheffe's test) between coda contexts, and the two neighboring stimuli differed too although not so significantly (see Fig. 1).

In order to assess the magnitude of the boundary shift, the location of the category boundary (i.e. the vowel duration corresponding to a 50% 'short/long' labeling) was computed individually for both curves of every listener using the procedure described in [5]. The difference between the mean boundary location in the voiced-coda context (100.89 ms) and the mean boundary location in the voiceless-coda context (96.95 ms) was approximately 3.93

ms. The boundary location scores were subjected to repeated-measures ANOVA which showed a significant main effect of coda context (F[1, 53]=20.751, p=.00003).

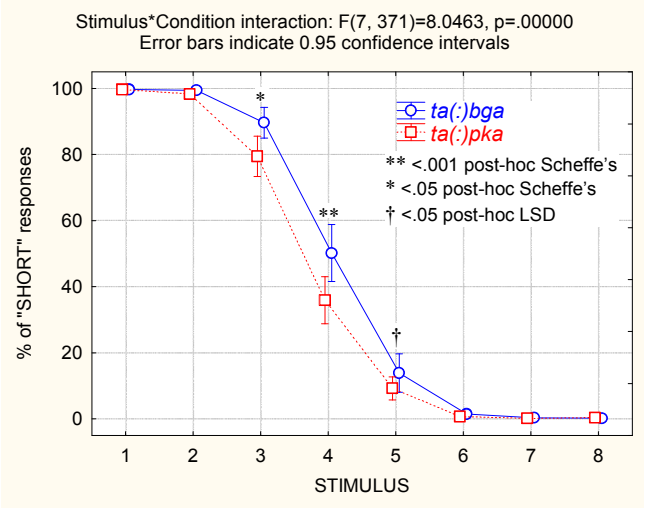

Fig. 1 Native Czech short/long vowel categorization. Average identification curves for the voiceless (dotted) and voiced (solid) coda context. Repeated-measures ANOVA found a significant interaction between vowel duration and coda voicing. Post-hoc tests showed differences in responses to near-boundary vowel durations.

When response patterns were considered within individuals, not all listeners showed a boundary shift similar to the average one. Individual response patterns were sorted into three classes: (1) a shift in the expected direction (i.e. the one the average identification curves show), (2) a shift in the opposite direction, and (3) no clear shift (cases where there was either very little difference between curves or it was a difference in steepness rather than in location of the cut-off point). There were 33 expected-direction cases, 9 opposite-direction cases, and 12 no-clear-shift cases.

It was necessary to consider the order in which listeners completed the two tests in this experiment as a potential factor at play. Within the expected-direction group, the number of subjects who took the *tabga* test first was higher (19) than of those who took the *tapka* test first (14). However a χ^2 test of the whole 'response pattern' by 'order of tests' distribution (χ^2=.352) did not eliminate the null hypothesis (even at α=.1) and therefore the character of the boundary shift probably did not depend on the order of the tests (contingency coefficient C<.081).

3.2 Results for non-native listeners

No statistically significant differences between the two tests in the non-native listeners' group were found. The average identification functions were relatively categorical and thus resembled the average curves of the native listeners in the steepness and they did so as well in the approximate location of the category boundary (102.97 in the voiced consonantal context and 103.6 in the voiceless context). Importantly, the location of the short/long boundary did not differ between the two tests: repeated-measures ANOVA showed no significant interaction between coda voicing and vowel duration (p>.6).

It was possible that because all subjects participated in both tests a learning effect may have occurred. To factor out the possibility that such a learning effect could be distorting the data, only results of the first test taken by each non-native listener were analyzed next. Since the order in which the tests were completed was counterbalanced, there were now 8 curves for the voiced context and 8 curves for the voiceless context to be examined. For each curve, the 50% crossover point was again computed. Even now, there were no significant differences between consonantal contexts in the boundary scores (t-test, α=.1).

4 Discussion

The first aim of this study was to obtain baseline information about Czech perception of vowel quantity contrasts in different consonantal voicing contexts. The specific question was whether or not vowel length perception is affected by the voicing of the following tautosyllabic obstruent (the presumption being that in Czech vowel duration varies as a function of the voicing context and that listeners are sensitive to this variation).

The present findings about native Czech perception suggest that listeners indeed adjust the location of the short/long vowel boundary so that a vowel needs to last slightly longer when followed by a voiced coda to be perceived as long than it does when followed by a voiceless coda (the provision being that not all individuals in this study followed this pattern). Although the size of the average category boundary shift was very small (3.93 ms), listeners seem to be very consistent in adjusting their perception of phonemic vowel quantity, considering the high significance of the Vowel duration × Coda voicing interaction and of the main effect of Coda voicing on the boundary location scores.

The finding that in Czech a quantitatively ambiguous vowel is more likely to be perceived as short before a (phonetically) voiced coda obstruent than before a voiceless coda obstruent provides indirect support to the hypothesis that in Czech (like in many different languages, see e.g. [7]) a voiced obstruent causes a lengthening (or alternatively, that a voiceless obstruent causes clipping) of the nucleus of the syllable it closes.

The second and primary purpose of this study was to examine if allophonic use of vowel duration in the L1 hinders accurate perception of L2 phonemic vowel length contrasts. On the basis of earlier studies (e.g. [2, 4, 6]) transfer of L1 perceptual strategies to the L2 appeared to be likely. In the present study it was hypothesized that such transfer would cause English learners of Czech to shift the boundary between phonologically short and long vowels considerably depending on the voicing of the coda consonant. Nevertheless, no difference between short/long vowel categorization before a voiced coda and a voiceless coda was found for the native English listeners. It is true that in this study the number of non-native listeners was relatively small. However I assume that if transfer took place it should result in a boundary shift observable even in small samples (given the relatively large variation in vowel duration due to coda voicing in English [8]).

The absence of any biases in the non-native responses attributable to the influence of L1 is interpreted as an indication that the L2 listeners in the present study successfully dissociated vowel duration from coda-obstruent voicing which it marks in English and thus

redefined vowel duration as a cue for L2 vowel quantity. Segmental duration has been hypothesized [16] to be a readily available aspect of the speech signal and it is possible that duration is a salient cue even for speakers of languages without phonological quantity contrasts.

Acknowledgments

Many thanks are extended to Kateřina Chládková (a student at Palacký University), who helped to collect data and commented on an earlier draft of this paper, and to Šárka Šimáčková (Palacký University), who offered advice on the design of the experiment.

References

[1] C. T. Best, "A direct realist view of cross-language speech perception", In: *Speech perception and linguistic experience: issues in cross-language research*, (W. Strange, ed.), Baltimore: York Press, 171-203 (1995)

[2] R. McAllister, J. E. Flege, T. Piske, "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian", *J. Phonetics* 30, 229-258 (2002)

[3] J. E. Flege, "Second language speech learning: theory, findings, and problems", In: *Speech perception and linguistic experience: issues in cross-language research*, (W. Strange, ed.), Baltimore: York Press, 233-277 (1995)

[4] K. Tajima, H. Kato, A. Rothwell, R. Akahane-Yamada, K. G. Munhall, "Training English listeners to perceive phonemic length contrasts in Japanese", *J. Acoust. Soc. Am.* 123, 397-413 (2008)

[5] S. Ylinen, A. Shestakova, P. Alku, M. Huotilainen, "The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish", *Language and Speech* 48, 313-338 (2005)

[6] Y. Minagawa-Kawai, K. Maekawa, S. Kiritani, "Effects of pitch accent and syllable position in identifying Japanese long and short vowels: comparison of English and Korean speakers", *Journal of the Phonetic Society of Japan* 6, 88-97 (2002)

[7] M. Chen, "Vowel length variation as a function of the voicing of the consonant environment", *Phonetica* 22, 129-159 (1970)

[8] D. H. Klatt, "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", *J. Acoust. Soc. Am.* 59, 1208-1221 (1976)

[9] T. Walsh, F. Parker, "Vowel length and 'voicing' in a following consonant", *J. Phonetics* 9, 305-308 (1981)

[10] I. Lehiste, *Suprasegmentals*, Cambridge, MA: MIT Press (1970)

[11] N. Warner, A. Jongman, J. Sereno, R. R. Kemps, "Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch", *J. Phonetics* 32, 251–276 (2004)

[12] P. Machač, R. Skarnitzl, "Temporal compensation in Czech?", *Proc. 16th ICPhS* Saarbrücken, 537-540 (2007)

[13] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese", *J. Phonetics* 32, 565-589 (2004)

[14] Z. Palková, *Fonetika a fonologie češtiny s obecným úvodem do problematiky oboru*, Prague: Karolinum (1994)

[15] T. Scheer, "On the status of word-initial clusters in Slavic (and elsewhere)", In: *Formal Approaches to Slavic Linguistics 15: the Toronto Meeting*, (R. Compton, M. Goledzinowska, U. Savchenko, eds.), Ann Arbor: Michigan Slavic Publications, 346-364 (2007)

[16] O.-S. Bohn, "Cross-language speech perception in adults: first language transfer doesn't tell it all", In: *Speech perception and linguistic experience: issues in cross-language research*, (W. Strange, ed.), Baltimore: York Press, 279-304 (1995)