"Engela's Eshes": Cross-linguistic perception and production of English [æ] and [ɛ] by Czech EFL learners trained in phonetics

Šárka Šimáčková

Palacký University Olomouc, Czech Republic

E-mail: simackov@ffnw.upol.cz

ABSTRACT

Studies of cross-language speech perception suggest that nonnative listeners make a greater use of duration in identifying and discriminating English vowels than native listeners do. Our first experiment investigates the role of duration as a cue in crosslinguistic identification of vowels [æ] and [ɛ] in natural and edited tokens of monosyllabic words. The subjects, Czech EFL learners, in whose L1 vowel length is phonemic, rely strongly on increased duration in identifying [æ] and cope poorly with temporal variability due to segmental context and computer editing. A complementary production experiment tests the efficiency of duration as a marker of [æ]-[ɛ] contrast in Czech pronunciation of English. The learners produce a difference that is statistically significant but relatively small.

1. INTRODUCTION

The present goal is to investigate the role of duration in identifying English non-high front vowels [æ] and [ɛ] by Czech EFL learners and its role in Czech realization of the [æ]-[ɛ] contrast. The study was conducted in the context of a Czech higher education institution, with subjects who are prospective language professionals (language teachers, translators/interpreters) and who are thus expected to achieve higher proficiency in English than an average Czech EFL learner. Their training is both more intensive and extensive than it would be in other forms of language education, with the curriculum involving separate language modules such as English phonetics.

The phonetics course combines theory with practical training. The latter targets primarily receptive skills. Although, those skills improve over time, there are several persistent difficulties. Among the aspects of production and perception that prove most resistant to training is differentiation between the vowels [æ] and [ɛ]. Transcriptions such as [ɛŋdʒələz ɛʃɪz] for *Angela's Ashes* or [ɛləbɑmə] / [ɛləbɛːmə] for *Alabama* are produced even at the end of the course.

Of course, Czech learners of English are not the only EFL population struggling with the English low front vowel. For example, speakers of German, Portuguese, Spanish, Korean, and Arabic also find it difficult to discriminate [æ] from neighboring vowels, [1]. Substituting [ɛ] for [æ] by Czechs appears to be a straightforward example of *perceptual assimilation* – two contrasting L2 segments are absorbed by a single L1 category. One of the vowels is a better and the other a poorer example of the L1 category, [2]. When transcribing, learners trained in IPA represent both L2 vowels with an epsilon. For them, the symbol represents the vowel quality recognized as their L1 vowel [e]. Czech [e] is not differentiated from L2 [ɛ] which is not differentiated from L2 [æ].

The [æ]-[ɛ] misidentifications, however, are not unidirectional. The same learners also transcribe [əmæɹɪkə] for *America* or [ʃæfs] for *chefs*. This gives the impression that learners' choice between epsilon and ash is random. They are instructed about the existence of ash-epsilon contrast in English and they suppose that ash must be appropriate sometimes. And yet, this impression of randomness disagrees with the fact that transcriptions from different learners converge on epsilon↔ash misidentifications. The value of the vowel they chose in a specific case is predictable. Inspection of tokens of [ɛ] mistranscribed as [æ] shows that they typically occur in the peak of a tonic syllable or in the final syllable of a tone unit, and as such are realized as longer. This study tests whether vowel duration is the predictor of learners' choices and whether duration serves to distinguish epsilon and ash also in learners' productions. Besides incidental observations of subjects' transcription habits, there are other reasons why duration is expected to play an important role:

a) *Temporal properties of* [æ]. Though classified as a lax vowel phonologically, [æ] is longer than [ɛ] and other lax vowels, especially when followed by a voiced obstruent.

b) *Contrastive function of duration in Czech.* Listeners sensitized to duration by L1 can be predicted to use the temporal difference between epsilon and ash as a cue in identification of the vowels.

c) *Results of previous studies.* EFL learners from disparate L1 backgrounds have been shown to rely on duration rather than spectral quality in identifying English vowels, e.g. [3-5].

ISBN 1-876346-48-5

In what follows, [æ] and [ɛ] are considered in relation to the other front vowels. English vowel system contrasts four levels of height - /i, ɪ, ɛ, æ/. Duration is an inherent property of a vowel. It varies allophonically according to the voicing of the following obstruent (as well as phonotactics, word-stress, and prosodic context).

The Czech *i-e-a-o-u* vowel inventory differentiates three levels of height. Vowel quantity is phonemic. Four front vowels defined by two levels of height and two values of length are contrasted: /i, iː, e, eː/. The high front pair transcribed phonetically as [iː, ɪ], is less similar in quality than the other temporal pairs. Obstruent voicing does not condition variation in vowel duration, phonologically voiced obstruents are devoiced word finally.

Contrastive analysis of Czech and English vowels cannot predict whether [æ] should assimilate along the high-low dimension or along the front-back dimension. Assimilation to Czech [a] is also possible, as is shown by many older loan words, though it has been seldom observed in learners' performance. For this reason, the production task also includes low back vowels.

2. PERCEPTION: METHODOLOGY

Six American women from Nebraska in their early twenties were recorded saying twenty English bVd/bVt words in the carrier sentence *'Say__ please.'*: ***bed/bet, bad/bat****, bid/bit, bead/beat, bud/but, boot/booed, bawd/bought, Bob/bop, bayed/bate, bode/boat.*

From each talker, six tokens of each word were collected, low-pass filtered at 4,8 kHz and digitized at 10 kHz sampling rate. F1 and F2 of [i, ɪ, ɛ, æ] at the midpoint of the vowel were estimated from LPC (14 coefficients) and converted into Barks. Duration of the four vowels was measured. The six Americans differentiated [æ]-[ɛ] in terms of height, with [æ] being on average 1,4 Bark lower than [ɛ] (squares in Fig. 1). [æ] was also 1,5 times longer than [ɛ]. From this pool, words *bad, bat, bed, bet* of two speakers (K and S diamonds in Fig. 1) were selected to serve as the target stimuli and their other words as distractors. The selection was based on comparison of token quality - the final consonant was not realized as a glottal stop, the vowels did not end in a creaky voice and were monophthongal throughout their duration. The target words *-bad/bed/bat/bet-* were used once in their natural form and once with edited duration. Distractors were not edited. Table 1 gives spectral information about the four target tokens from each talker. Editing involved altering duration of [æ] and [ɛ]. [æ] of *bad/bat* was shortened to [ɛ] of *bed/bet* by deleting pitch pulses from the steady portion of the vowel. [ɛ] of *bed/bet* was extended by reduplicating pitch pulses in the steady portion of the vowel. In this way the duration of the opposite token was approximated as closely as possible (Fig. 2).

Combining 4 targets (*bad/bat/bed/bet*), 2 talkers (K/S), and 2 conditions (natural/edited) yielded 16 test words that were complemented by 16 distractors. Eight digitized strings of 10 randomized words (2 test words, 8 distractors) were created with 10 ms separating the words. The 8 strings were played over the headphones to the subjects sitting at computers. 30s pause separated one string from the next, each was played only once. Upon hearing a word subjects immediately transcribed the vowel in IPA. Two practice sessions preceded the experiment.

36 Czech female college students aged 19-21 years majoring in English and 6 American female college students from three Mid-West colleges participated.

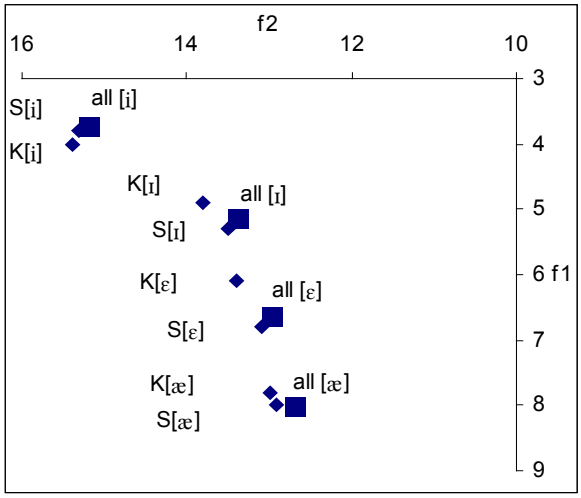

Figure 1. American [i, ɪ, ɛ, æ]. *All*: 6 talkers averaged. *K, S*: the stimulus talkers (6 tokens averaged).

	F1 (Hz)		F2 (Hz)	
	K	S	K	S
bat	917	939	2066	1987
bad	904	928	2055	2018
bet	689	740	2079	2023
bed	619	749	2137	2124

Table 1. Mid-point formant frequencies of target vowels.

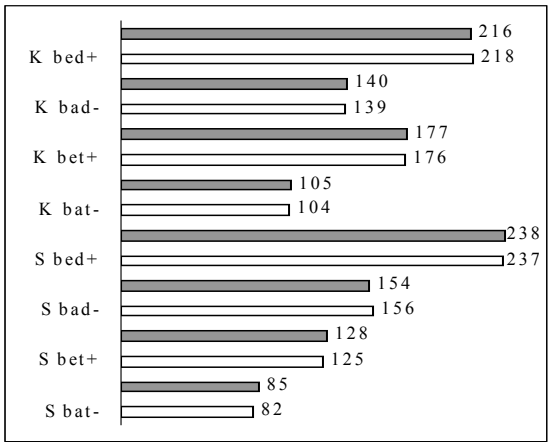

Figure 2. Duration of the target words in ms. Light bars are edited targets ("+" is extending, "-" is shortening).

3. PERCEPTION: RESULTS

Native speakers of English avoided misidentifications of the natural targets and performed near the ceiling on the edited targets. Czech listeners identified *bad/bat* vowel as [æ] or [ɛ] and less frequently as [ɑ]; *bed/bet* vowel was identified as [ɛ] or [æ], and also as [ɪ].

Listeners' responses were tallied. For both [ɛ] and [æ] in each condition and voicing context, frequency of each response category was counted. A response that corresponded to the target was labeled as a match, a response that did not corresponded to the target as a mismatch. Two way Chi-square was used to check if frequencies of matches and mismatches changed in response vowel quality, consonant voicing, condition, and talker identity.

In quality, both English [æ] and [ɛ] are likely to assimilate perceptually to the Czech vowel [e]. This should enhance the reliability of identification of [ɛ], which is a *similar* vowel, and reduce reliability of identification of [æ] – a *new* vowel assimilated to the same category. This expectation is not supported, natural [ɛ] is just as likely to be mistranscribed as natural [ae] (χ^2=1,11, p>0,05).

If duration determines recognition of a vowel sound as [æ] or [ɛ], token's duration will be correlated with the number of times the token is identified as [æ] ([ɛ]). This is confirmed, r= 0,9 for [æ] and r = – 0.88 for [ɛ].

If duration cues learner's response, voicing context will generate differential response to natural targets: [æ] will be identified less consistently in *bat* than in *bad* and [ɛ] will be identified less consistently in *bed* than in *bet*. Percentages of matches and mismatches for each target are shown in Fig. 3 (light bars). The prediction is borne out in the case of [æ]. The rate of correct identification of [æ] is significantly higher for *bad* than for *bat* (χ^2=15,84, p<0,0001). Only in *bad* is [æ] mistranscribed as long tense [ɑ]. The prediction is only partially supported in the case of [ɛ]. The number of matches is not different for *bet* and *bed* but only in *bed* is the vowel mistranscribed as [æ], in *bet* it is mistranscribed as short lax [ɪ].

If Czech perception of [æ]-[ɛ] contrast is quantity-driven, then editing duration of the target vowels will be detrimental to their correct identification. The gray bars in Fig. 3 show the pattern of responses to stimuli whose duration was manipulated (*bad¯/bat¯* indicate a reduced token, *bed+/bet+* indicate an extended token). The difference in frequency of matches between natural and edited stimuli is significant: p<0,05 for [ɛ] and p<0,001for [æ] – evidently, [æ] is more affected by editing. The percentage of mismatches is 4 times higher when duration of [æ] in *bad* is shortened to that of [ɛ] in *bed*. While [æ] in natural *bat* is identified at the level of chance, in edited *bat* the vowel is quite consistently perceived as [ɛ]. Vowel [ɛ] is more robust in the face of lengthening. Extending duration of [ɛ] in *bet* to that of [æ] in *bat* is not sufficient to cause a dramatic change in the identification rate although there is a 9% drop in matches. On the other hand, lengthening of [ɛ] in *bed* to [æ] in *bad*, brings identification down to the level of chance. Notice though that even when *bed* is equalized with *bad* in terms of duration, the rate of [æ] responses is 34% higher for *bad* than for *bed+* (χ^2=18,4, p<0,0001). The choice of [æ] is 21% higher for *bat* than for *bet+* but the difference is not significant (χ^2=6,35, p<0,05). When the rate of [ɛ] responses is counted for *bed* and *bad¯*, there is a 16% difference in favor of the natural token but the difference is not significant (χ^2=3,74, p>0,05). In the case of *bet* and *bat¯*, the frequency of [ɛ] responses is very close. Shortening of [æ] in *bad/bat* produced significantly more misidentifications than lengthening of [ɛ] in *bed/bet* (χ^2=15,21, p<0,0001). Duration is more important for identification of the new vowel [æ].

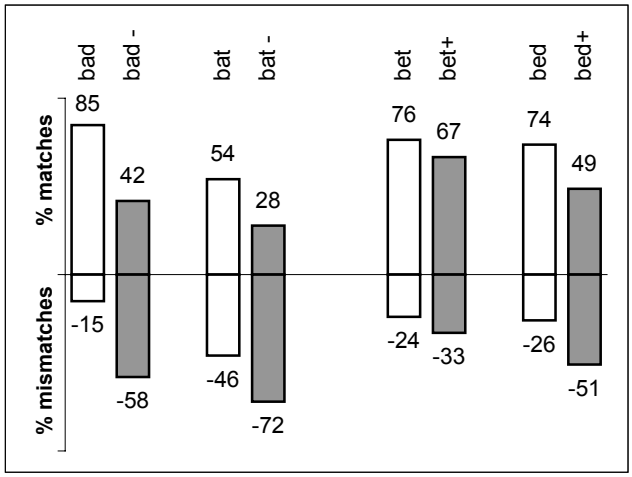

Figure 3. Percentage of matches (positive numbers) and mismatches (negative numbers).

The subjects responded to both talkers in a similar fashion. The only talker-related difference concerns responses to the edited [æ]. Recognition of [æ] in *bad¯/bat¯* is significantly poorer for K than for than S (p<0,01). The key to such differential response does not seem to be either in the mid-vowel spectral quality or in its quantity.

4. PRODUCTION: METHODOLOGY

Nonnative production data were elicited from the same 36 Czech women who participated in the perception experiment. Subjects were recorded saying 20 English *bVd/bVt* in the carrier sentence *'Say__ please'*. Six tokens of each word were recorded. These data are compared to the productions elicited from the six Americans described in section 2. Czech subjects also produced six tokens of 12 Czech monosyllabic words contrasting long and short vowels so that Czech-English as well as English-EFL

ISBN 1-876346-48-5

comparisons could be drawn. The Czech set included: ***mez/pes***, ***vézt/vést***, *miz/mis, míz/mís, maz/mas, máz/máš.*

T-tests were used to compare F1 and F2 of pairs of vowels. Significance of duration differences was tested by a 2 factor Anova (Vowel X Voicing) separately for the English, EFL, and Czech data.

5. PRODUCTION: RESULTS

Crosslinguistic comparisons of vowel quality were made first. With the exception of [i], which has a significantly higher F2 than [ɪ], neighboring front vowels of the six Americans are separated in terms of F1. Vowel [æ] is as front as [ɛ] but it is lower. Each of the Czech front vowels [iː], [ɪ], and [e, eː] differs significantly from the next one both in F1 and in F2. Czech [e, eː] is 1,1 Bark higher than English [ɛ].

Czech pronunciation of English [æ] differs substantially from Czech [a, aː] in having significantly higher F2, and it is closer in quality to Czech [e, eː]. However, learners' [æ] and [ɛ] do not simply assimilate to the same L1 category. In height, both vowels are statistically different from Czech [e, eː] and they are different from each other. This may be an embryonic stage of [æ]-[ɛ] contrast formation. However, [æ]-[ɛ] difference in Czech EFL is considerably smaller than in native speech (0,4 compared to 1,4 Barks) and, critically, nonnative [æ] overlaps with American [ɛ]. Nonnative [ɛ] is between Czech [e] and American [ɛ].

In terms of duration, English [i, ɪ, ɛ, æ] and the respective Czech equivalents [i, ɪ, e, eː] are more alike when a voiceless obstruent follows. Compare the target pairs E[ɛ, æ] and C[e, eː] in Fig. 4. In the voiceless context, duration of English [æ] corresponds to that of Czech long [eː]; in the voiced context, [æ] is 60 ms longer than its Czech counterpart. Similarly, English [ɛ] is 60 ms longer than Czech short [e] before a voiced obstruent and only 21 ms before a voiceless one.

Czech EFL learners produced a duration difference between the English vowels. Their [æ] is significantly longer than [ɛ] both in the voiced and voiceless context (30 and 29 ms respectively). However, the [æ]-[ɛ] difference is small compared to the speech of Americans, especially for *bad* - *bed* (86 ms). Czechs also produced somewhat longer vowels in *bed/bad* compared to *bet/bat.* But the difference of 24 and 23 ms respectively was too small for voicing to be a significant factor. As is evident from Fig. 4, it is the vowel [æ] in *bad* that differs most in the production of the Czechs (CE[æ]) and the native speakers (E[æ]).

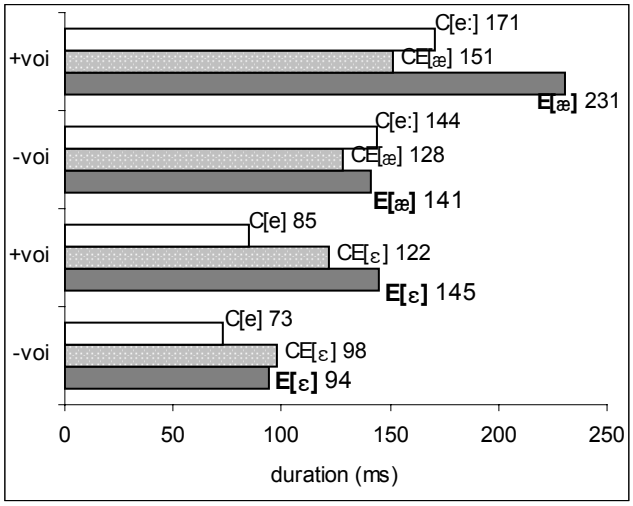

Figure 4. Duration of [ɛ, æ] and [e, eː] in Am. English (E), Czech English (CE), and Czech (C).

6. CONCLUSIONS

The EFL learners relied on duration as a cue in identifying English [æ]. The long [æ] in natural *bad* was recognized most reliably of all targets and [ɛ] was misheard as [æ] only when prolonged in *bed*. Production did not match perception. Learners' [æ] in *bad* approximated native English [æ] least well. Thus, in words such as *bad*, *cab*, or *rag* Czechs produce [æ] that is too short, spectrally overlaps with [ɛ], and that is typically followed by a completely devoiced obstruent. A native listener must rely on the context not to interpret them as *bat, cap*, or *wreck*.

REFERENCES

[1] J. E. Flege, "Second language speech theory, findings and problems," in *Speech perception and linguistic experience*, W. Strange, Ed., pp. 233-277. Baltimore: York Press, Strange, 1995.

[2] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience*, W. Strange, Ed., pp. 171-203. Baltimore: York Press, 1995.

[3] O.-S. Bohn, and J. E. Flege, "Interlingual identification and the role of foreign language experience in L2 vowel perception," *Applied Psycholinguistics* 11, pp. 303-328, 1990.

[4] O.-S Bohn, "Cross language speech perception in adults first language transfer doesn't tell it all," in *Speech perception and linguistic experience*, W. Strange, Ed., pp. 279-304. Baltimore: York Press, 1995.

[5] P. Escudero, *Developmental patterns in the adult L2 acquisition of new contrasts: the acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speaker,* University of Edinburgh, 2000.