The open front vowel /æ/ in the production and perception of Czech students of English

Pavel Šturm, Radek Skarnitzl

Institute of Phonetics, Charles University in Prague, Czech Republic

sturmp@seznam.cz, radek.skarnitzl@ff.cuni.cz

Abstract

This study addresses the acquisition of the English open front vowel by Czech learners of English, who are known to experience difficulties in both its production and perception. Secondary school students and university students of English judged the acceptability of the open front vowel as pronounced by other Czech learners of English. Their evaluations were plotted against acoustic measurements (F1, F2, and vowel duration) and linguistically relevant variables. The evaluations varied as a function of F1 and L2 experience. The experienced subjects perceived the vowel more accurately and consistently than did the relatively inexperienced subjects.

Index Terms: L2 acquisition, vowel perception, foreign accent, language experience, Czech English

1. Introduction

Adult learners of a second language (L2) usually attain a native-like pronunciation of that language only exceptionally. A number of studies (discussed in [1], [2], [3]) have shown that when L2 learners begin learning a new language after a certain age, usually around the onset of puberty or even much earlier, their capacity for learning new phonetic contrasts is limited, and their pronunciation shows a perceptible foreign accent. This “critical period” hypothesis is nevertheless a point of disagreement among researchers, and some rather suggest a “sensitive period”, during which the acquisition of linguistic abilities is successful, and after which it may not be so regular or complete [2]. In any case, it is well known that unlike infants, who can differentiate nearly all phonetic contrasts [3], adults perceive non-native speech sounds in a way biased by their first-language (L1) phonemic categories. Some contrasts are easily recognizable and reproducible by L2 learners, while other contrasts may pose problems.

There are various models aiming to describe and explain the acquisition of L2 and the perception of L2 contrasts; for a good summary see [1], [4]. All these models correctly predict that, at least in the initial stages of L2 acquisition, some L2 phones will be pronounced differently by non-native speakers, who are liable to interpret the non-native sounds through the native categories [1], [4], [5]. The deviations of L2 learners from the native speakers’ phonetic norms often lead to a perceived foreign accent, which is not restricted to the segmental level only – suprasegmental and subsegmental differences have also been shown to contribute significantly [6].

The present paper investigates one of the features that are commonly associated with Czech English. Leaving the differences in consonant inventories aside, Czech learners must handle a completely different system of vowels. The vocalic inventory of Czech (or, specifically, its monophthongs) consists of five short and five long vowels. The short vowels have identical quality as the long vowels, with the exception of /ɪ/ and /iː/, which differ more in quality than in quantity [7]. Figure 1 compares the Czech and English vowel systems. The values for English are taken from [8] (5 speakers, 4 tokens of each vowel); the age group of 35–40 was preferred because it is very unlikely that the shift which the vowel /æ/ is currently undergoing in British English would be reflected in Czech English. The values for Czech are taken from [9] (75 speakers, 8 tokens of each short vowel); the formants of the long /iː/ have been taken from [7].

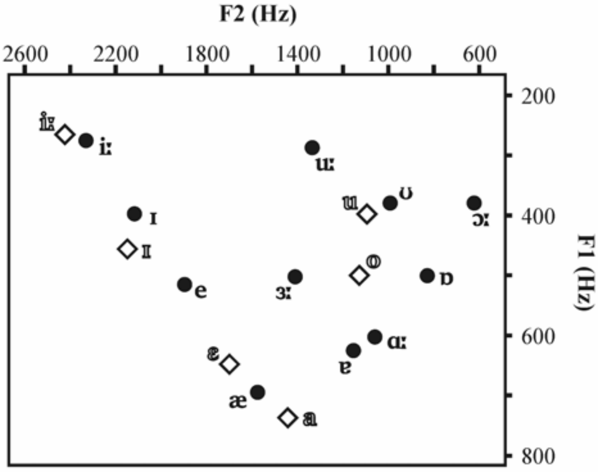

Figure 1: *Formant values of monophthong vowels in Czech (empty diamonds) and English (filled dots).*

It is clear that the differences illustrated in Figure 1 present several potential problems for Czech learners of English, especially in the area of open vowels: while there is only one vowel, /a/, in Czech, there are three vowels in English, /æ/, /ɐ/, and /ɑː/ (note that in monolingual materials, /ɐ/ is traditionally transcribed as /ʌ/). This paper focuses on the acquisition of the vowel /æ/, which is frequently mispronounced and may even be a source of misunderstanding.

2. Method

Ten male native Czech students of English, aged between 19 and 23, were recorded at the beginning of their university studies. After sufficient preparation they were instructed to read, fluently and naturally, a series of BBC news-bulletins, which yielded 295 words with /æ/ in total; words containing /æ/ with non-English proper nouns had been discarded. Another 51 words were discarded, mainly because of an obvious phoneme substitution (e.g. /e͡ɪ/, /ɐ/). Thus the analysis was based on 244 realizations of the vowel /æ/ (example words: “statues”, “banking”, “cabinet”, “animals”). The signal was segmented with respect to the waveform, spectral properties and formant patterns, and auditory inspection (for more detail regarding the segmentation guidelines, see [10]). In duration measurements, the start point after aspiration or glottalization was fixed as the onset of modal voicing. F1 and

F2 were measured at 7 equidistant points in the middle third of the vowel using the Burg method implemented in the Praat software [11] and then averaged. Default parameters were used except for Maximum number of formants (3) and Maximum formant (3000 Hz). In order to reduce the error of formant extraction, 20 % of the lowest and highest values of both F1 and F2 were manually checked and corrected by the first author; most of the errors were caused by coarticulatory nasalization of /æ/.

Out of these, 44 words were selected to be used in the perceptual testing according to the following criteria: (a) F1 values, (b) speaker identity, and (c) absence of distortion. The items were arranged according to F1 (which more or less correlates with the degree of openness) and divided into thirds. The three groups were meant to ensure that there were both extremes included in the test items, i.e. items with low F1 (approaching the more closed variant of Czech /ɛ/) and items with relatively high F1 (a more prototypical Czech /ɛ/ or even approaching the English /æ/; see the Introduction for details); the middle third comprises items with intermediate values. The number of items for the perception test was balanced for speaker. Finally, only such words were selected which lacked any marked distortion which might divert the attention from the target vowel (mispronunciation, but also the absence of aspiration of fortis plosives, etc.) and potentially lead to being affected in the judgment by a different phenomenon. Six items were selected for repetition so that intra-listener consistency might be checked. Two more items, not analyzed in the listening test, were added for initial training.

The perception test was administered to two groups of students; none of the speakers mentioned above took part in the listening test. Group A consisted of 43 Czech students (aged 19 to 23; 35 female + 8 male) of English studies at the Faculty of Arts who had already been formally acquainted with the English vocalic system in an introductory phonetics course, while Group B comprised 31 secondary school students (aged 17 to 18; 23 female + 8 male) not formally acquainted with the English vocalic system (apart from the standard secondary-school curriculum). The testing was conducted via high-quality loudspeakers in a sound-treated classroom adjusted for listening experiments. The listeners from Group A were divided into five groups, with each group listening to the items in a different randomized order; there were two randomizations for Group B. Each test item consisted of three repetitions of the word, followed by a desensitization passage of noise and synthesized tones. The total duration of the test was approximately 10 minutes.

The listeners were instructed to mark the degree of acceptability of the vowel /æ/, with reference to the standard British (RP) realization, disregarding any other factors. They were offered a three-point scale with the following labels: a) unsatisfactory, b) acceptable, and c) excellent. The three categories received commentary from the administrator of the test, *unsatisfactory* being described as “not good enough, should be better”, *excellent* as “approaching native-like quality”, and *acceptable* as “satisfactory”. The decision to use a three-point scale was taken for the following reasons. A binary choice was regarded as too crude; since we were especially interested in the extremes of openness (how close/open the vowel must be to be perceived as inadequate), we decided for the three-point scale. Moreover, we believed that it would be most intuitive and therefore easy to mentally grasp for the students; as Southwood and Flege [6] have shown, the reliability of the selected scale is an important factor. Given this design, it is obvious that the judgments had to be analyzed as nonparametric data due to unequal steps between categories.

Frequencies in Hz were converted to an auditory scale, the Equivalent Rectangular Bandwidth (ERB), using Moore and Glasberg’s formula cited in [9]:

$$ERB = 11.17 \ln \left| \frac{f + 0.312}{f + 14.675} \right| + 43 \qquad (1)$$

where f is frequency (Hz). The ERB scale was chosen for formant normalization for these reasons: (1) although Lobanov’s z-score transformation came best at the evaluation of several normalization techniques [9], the current material (single-word extracts) is not sufficient for extrinsic normalization based on several vowels; (2) the ERB values are easier to compute; (3) vowel systems in literature are usually described in barks or ERBs, so comparison is easier.

3. Results

Figure 2 compares the evaluations for Group A (university students) and Group B (secondary-school students). Median scores for the two groups were 1.83 and 2.06, respectively; the distributions differed significantly (Mann-Whitney U = 767.5, $n_1 = n_2 = 50$, $p < 0.001$). Thus the experienced listeners (i.e., formally trained in English phonetics) were generally stricter in their evaluations than the relatively inexperienced listeners. As a group they used a wider range of rankings (a higher maximal mean score of an item), and their quartile range is larger and shifted into lower values.

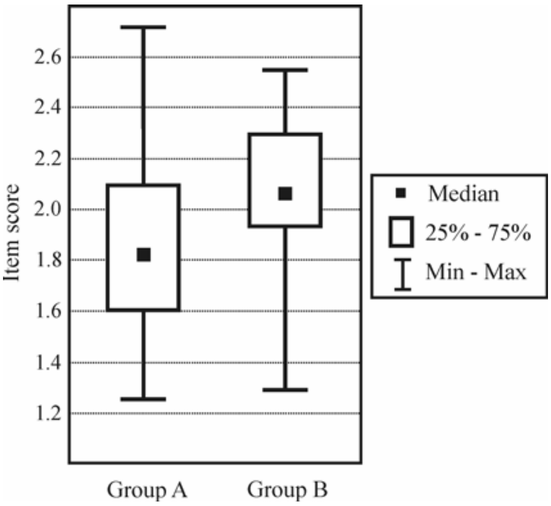

Figure 2: *Score values for Group A (experienced students) and Group B (inexperienced students).*

The two groups differ in the way they take into account the degree of openness. There is a strong correlation between F1 and item score for Group A (Spearman’s r = 0.71; p < 0.001), but no significant correlation for Group B (r = 0.12; p = 0.40). Correlation between the formant ratio (F2/F1) and item score is also strong for Group A (r = -0.60; p < 0.001), but there is no significant correlation between F2 and item score in either group.

Duration correlates with item score only weakly, even if items with /æ/ in the initial position are excluded from the analysis (the presence of glottal stop [ʔ] has the effect of shortening the following vowel, as shown in [12]; the results would therefore be less accurate). Within the group of items in which /æ/ is followed by a nasal consonant, there is quite a strong correlation between duration and item score for Group

A ($r = 0.57$; $p < 0.01$), but virtually no correlation for Group B. However, correlation is weak (both for Group A and for Group B) within the items in non-nasal contexts. A study dedicated specifically to the duration domain, with controlled contextual variables, would be needed for a better insight into the problem.

Figures 3 and 4 plot, for groups A and B respectively, the F1 and F2 frequencies of all target vowels against each other on an x-y scatter graph, and item scores for the given group are represented by different colours (black vs. grey). The mean score of 2.0, chosen as the boundary dividing good and poor instances of the vowel, is arbitrary but believed to be an adequate solution. As we can see in Figure 3, the situation in Group A is not at all clear-cut; the general tendency based on the correlation between F1 and acceptability is present, but with several exceptions. There are nine items in the vicinity of [æ] that have been given above-average scores, but also three (or even six) items with below-average scores; furthermore, the region of [e] accommodates four items with above-average scores, which runs contrary to our expectations. However, Group B (shown in Figure 4) is even less predictable in the evaluation, as the scores are scattered all over the plane without any distinct clusters. Moreover, students from this group were four times inconsistent in the evaluation of repeated items, and therefore the items in question had to be marked as both above- and under-average (empty circles). Comparing these results with the correlation data above, the evaluations by secondary school students seem to be rather random, not reflecting in any observable way the degree of openness, while there is a tendency for university students, who had been trained in both the production and perception of the English vowels (both generally and specifically with a focus on the difficult vowel /æ/), to base their decisions about vowel acceptability on this parameter.

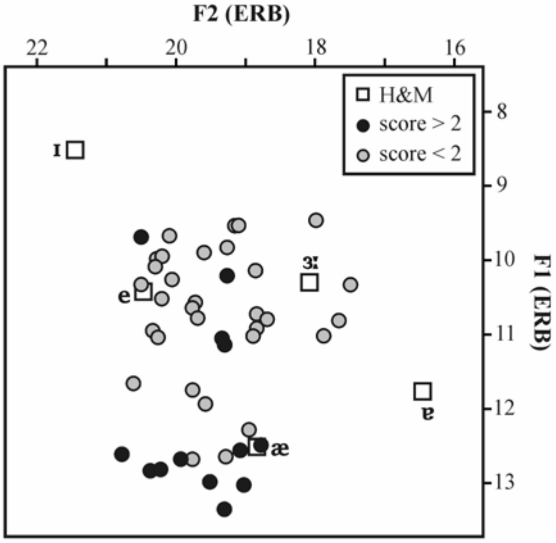

Figure 3: *F1~F2 plot for all items and Group A. Item scores are differentiated by colour. H&M are values taken from [8].*

A more thorough analysis of the items in the vicinity of [æ] was conducted with the aim to provide an explanation for the odd scores in that area for Group A (Figure 3); the other group was not analyzed because of its general inconsistency. First of all, the three items which have ideal values for /æ/ (F1 > 12 ERB) but which received lower evaluations do not deviate much: their scores are 1.9, 1.9, and 1.8; this holds true for the three items that are relatively more distant from the prototype of /æ/ as well (1.8, 1.9, but 1.5). Secondly, in five out of these six items /æ/ occurs in the initial position, preceded by the glottal stop [ʔ]: an atypical feature of native English speech, where various linking phenomena are more usual [13: 305f.], but typical of the speech of Czech speakers of English [14]. As the listeners were experienced with the English language, they may have regarded it as a feature of foreign accent and mark it accordingly. Lastly, in all these six items /æ/ was followed by a nasal consonant. As nasalization typically results in a weaker amplitude of the first formant, this might also influence the perception of the vowel and its evaluation. Since none of the items were judged by the first author to be pronounced in a markedly insufficient way (compared to the other items in question), speaker identity as a significant factor may be ruled out.

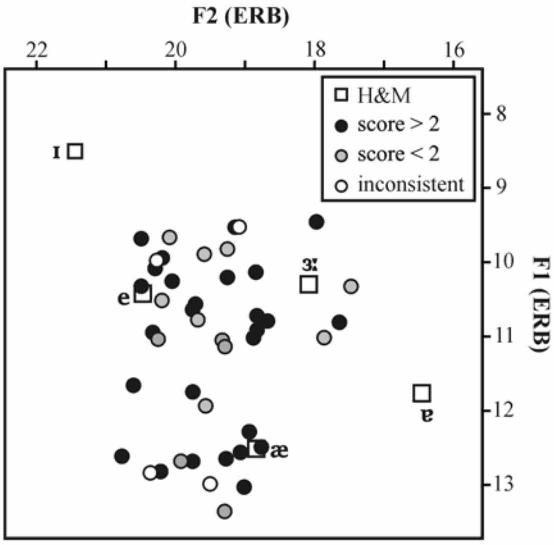

Figure 4: *F1~F2 plot for all items and Group B. See Figure 3 for explanation. Four repeated items that were evaluated inconsistently are marked with empty circles.*

As to the discrepancies in the upper part of Figure 3, one item can be explained quite easily. /æ/ in the word “statues” is surrounded by coronal sounds, and the listeners probably do not expect as open a pronunciation as in other contexts, and perceptually compensate for it. Unlike the vowel in isolation, which sounds like [ɛ], the word as a whole sounds adequate, and its score is accordingly very high (2.6). The three remaining above-average items in the area received only moderately high scores (2.1, 2.1, 2.2). Once again, it should be noted that the dividing point of 2.0 was chosen arbitrarily, and does not constitute a natural break.

As a next step, the items which had been assigned extreme scores (i.e., highest/lowest mean scores) were analyzed, since the extremes could show which parameters might play a prominent role in the perception of the acceptability of the vowel /æ/. Ten items with the lowest scores for Group A were selected and compared with the scores for Group B – secondary school students provided evaluations on average 0.6 points higher (i.e. milder). On the other hand, ten items with the highest scores for Group A were given evaluations on average 0.4 points lower by Group B. In other words, the extreme items behave according to the general pattern for the two groups as shown in Figure 2. Also, the items with extreme scores differ significantly in the mean values of F1: 10.0 ERB for the ten least acceptable items, as against 12.5 ERB for the ten most acceptable items (*t* (18) = -6.99; *p* < 0.001). There were no significant differences in F2. Duration was influenced by context (pre-fortis shortening), but the two groups of extreme items did not differ significantly in the duration domain.

Kruskal-Wallis ANOVAs were applied to the data but revealed no significant variations for speaker, the preceding context, the following context, the number of syllables in the word, or the position of the accented syllable in the word. This might partly be explained by the imbalance of the sample (e.g. there were 29 two-syllabic words, but only 2 four-syllabic words).

4. Discussion and conclusions

In accordance with the results reported in Flege et al. [5], it appears that experience with the L2 is an important factor in the perception of English vowels. University students who have undergone a formal instruction on the English vocalic system are, compared to secondary school students without such an instruction, (1) more consistent in their evaluations of the target vowel's acceptability, (2) stricter in their evaluations (have higher internal standards of the ideal pronunciation), and (3) more similar to native speakers with respect to the cues on which they base their decisions (the degree of openness) – Czech university students show a strong correlation between F1 and acceptability.

On the other hand, the present study does not confirm the findings from Strange [4] that the phonological status of vowel length in the L1 influences the perception of L2 vowels. Although vowel length is phonologically distinctive in Czech, and Czech students often lengthen the vowel /ɛ/ in order to imitate native speakers' pronunciation of /æ/, no significant correlation was found between the duration of /æ/ and its acceptability in either group. A possible explanation might be found in the type of material used. As our listeners were presented with single-word extracts from spoken sentences, the conditions were less favourable to perceptual normalization in the duration domain than would be the case if longer stretches of speech had been used.

An analysis of various contextual variables revealed that immediate context does not affect the evaluation significantly (unlike [4], where contextual variables have significant effects on the patterns of perceptual assimilation of vowels). Nor does the number of syllables in a word or the position of the accented syllable (and thus the vowel /æ/ in our material) appear to be a significant factor. Interestingly, however, there are differences in the dispersion of evaluations as a function of accent position and word length: when the accented syllable comes late in the word and/or the word is longer, variance is greater. This might indicate that listeners pay more attention to initial syllables than to final syllables; in Czech, it is the first syllable that is stressed.

It is obvious that acceptability ratings and other perceptual phenomena, such as judging the strength of foreign accent, do not have any explicit physical units on which listeners could base their decisions. The task is to a large degree subjective and impressionistic, and naturally, variability is quite high. Although the listeners were instructed explicitly to focus on the target vowel, they may have paid attention to other factors, albeit unconsciously; this effect could be reduced by using synthetic or re-synthesized speech. Most importantly, listeners may use different acoustic cues, or ascribe them different significance.

Experience with the target language turned out to be an important factor in our study, as the experienced listeners paid more attention to an acoustic cue (F1), and were more consistent, than the less experienced listeners. It is to be noted, though, that notwithstanding the degree of experience with the L2, Czech listeners' perceptual category for English /æ/ may still be, and is in fact likely to be, different from that of native listeners. Therefore it might be useful to include native English-speaking listeners as well, and compare their results with the results obtained in the current experiment. Another direction for future research lies in the domain of synthetic or re-synthesized speech. Formants in the material from natural speech might be manipulated, and the re-synthesized stimuli then presented to the listeners. The temporal domain is also worth a more thorough investigation, with carefully controlled contextual and prosodic variables.

5. Acknowledgements

This research was supported by a grant to the first author from the Faculty of Arts, Charles University in Prague.

6. References

[1] Lecumberri, M. L. G., Cooke, M. and Cutler, A., "Non-native speech perception in adverse conditions: A review", Speech Communication, 52, 2010: 864-886.

[2] Obler, L. K. and Hannigan, S., "Neurolinguistics of Second Language Acquisition and Use", in Ritchie, W. C. and Bhatia, T. K. [Eds], Handbook of Second Language Acquisition, 509-523, Academic Press, 1996.

[3] Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T. and Pruitt, J., "Early Speech Perception and Later Language Development: Implications for the 'Critical Period'", Language Learning and Development, 1, 2005: 237-264.

[4] Strange, W., "Levels of Abstraction in Characterizing Cross-Language Phonetic Similarity", in Proceedings of 14th ICPhS, San Francisco, 2513-2519, 1999.

[5] Flege, J. E., Bohn, O.-S. and Jang, S., "Effects of experience on non-native speakers' production and perception of English vowels", Journal of Phonetics, 25, 1997: 437-470.

[6] Southwood, M. H. and Flege, J. E., "Scaling foreign accent: direct magnitude estimation versus interval scaling", Clinical Linguistics and Phonetics, 13, 1999: 335-349.

[7] Podlipský, V. J., Skarnitzl, R. and Volín, J., "High Front Vowels in Czech: A Contrast in Quantity or Quality?", in Proceedings of Interspeech, Brighton, 132-135, 2009.

[8] Hawkins, S. and Midgley, J., "Formant frequencies of RP monophthongs in four age groups of speakers", Journal of the International Phonetic Association, 35, 2005: 184-199.

[9] Volín, J. and Studenovský, D., "Normalization of Czech Vowels from Continuous Read Texts", in Proceedings of 16th ICPhS, Saarbrücken, 185-190, 2007.

[10] Machač, P. and Skarnitzl, R., "Principles of Phonetic Segmentation", Epocha, Praha, 2009.

[11] Boersma, P. and Weenink, D., "Praat - Doing phonetics by computer (Version 5.1.31.)", 2010, accessed on April 4, 2010, from http://www.praat.org/.

[12] Volín, J. and Skarnitzl, R., "Temporal downtrends in Czech read speech", in Proceedings of Interspeech, Antwerpen, 442-445, 2007.

[13] Cruttenden, A., "Gimson's Pronunciation of English", Hodder Education, London, 2008.

[14] Bissiri, M. P., Lecumberri, M. L., Cooke, M. and Volín, J., "The role of word-initial glottal stops in recognizing English words", in Proceedings of Interspeech, Florence, 2011.