Vowel Accentedness in the Light of Internal and External Competence Assessment

Jan Volín, Tanja Kocjančič Antolík, Radek Skarnitzl, and Pavel Šturm

Abstract It is generally accepted that vowel quality contributes to the overall impression of accentedness of speech, and affects both intelligibility and comprehensibility. The vocalic oppositions maintain the functionality of the language and past research has identified beg × bag contrast as the most troublesome for Czech learners of English. In this study, we add new empirical data on this contrast, but importantly, we also compare it with pot × port contrast, which is to some extent analogical but also different in both phonological and phonetic sense. Moreover, we relate our findings to (a) internal (own) and (b) external (experienced teachers') assessments of Czech EFL learners' pronunciation competence. Speech samples of 38 Czech learners of English show that individuals differ in their manifestations of foreign accent and that both internal and external assessments reflect the feature of openness in front and back vocalic pairs even in general impression assessments (i.e., without any focus on vowels). This finding highlights the importance of the openness feature in teaching EFL to Czech learners. In addition, the study provides specific formant values that can be used as reference data for cross-linguistic comparisons in future research.

Keywords Intelligibility · Comprehensibility · Native English accents · Non-native English accents

J. Volín (✉) · T. K. Antolík · R. Skarnitzl · P. Šturm
Institute of Phonetics, Charles University, Prague, Czech Republic
e-mail: jan.volin@ff.cuni.cz

T. K. Antolík
e-mail: tanja.kocjancicantolik@ff.cuni.cz

R. Skarnitzl
e-mail: radek.skarnitzl@ff.cuni.cz

P. Šturm
e-mail: Pavel.Sturm@ff.cuni.cz

187

1 Introduction

Foreign-accented speech is a phenomenon of continuous interest throughout human history, and the past decades have provided a more rigorous methodological framework of its research. We no longer base our claims on anecdotal observations, and the space for speculative explanations has been restricted quite substantially. Instead, quantitative empirical research is supported with its clear hypothesising and hypothesis testing. Such research has already resulted in detailed categorisations of various aspects of speech that contribute to accentedness.

One of the aspects could be labelled *vocalic*, that is, entailing vowels. It is a universal feature of human languages that they contain vowels in their segmental inventories. There is no known language that would use only consonants in its spoken form. The acoustic foundation of vowels rests on the fact that the raw voicing sound passing through the supralaryngeal cavities of the vocal tract induces greater vibrations at specific frequency bands. These stronger frequencies are perceptually salient and they are called *formants*. The first and second formant, marked $F1$ and $F2$ respectively, account for most differences between vowels, which makes them an elegant and parsimonious set of parameters. Vowel formants can be visualized in a two-dimensional plot and, as acoustic dimensions, they are closely related to the corresponding articulatory dimensions, that is, vowel height and the frontness or backness of the tongue. The $F1$–$F2$ plots may thus display vocalic differences between languages, language varieties, as well as individual speakers.

A modern approach to vowel formants dates back for almost two centuries. Formant specifications are used to capture vowel qualities, whether we are interested in (a) the vocalic system of a given language (e.g., Beňuš, 2012, for Slovak; Deterding, 1997, for Standard British English; Hillenbrand et al., 1995, for American English; Skarnitzl & Volín, 2012, for Czech), (b) comparing varieties of a language (e.g., Escudero et al., 2009, for Brazilian vs. European Portuguese; Ferragne & Pellegrino, 2010, for varieties of British English; Fox & Jacewicz, 2009, for varieties of American English; Hawkins & Midgley, 2005, for age-related variation), or (c) analysing vowels in foreign-accented speech (e.g., Leppik et al., 2019; Munro et al., 2013; Šimáčková & Podlipský, 2018; Šturm & Skarnitzl, 2011).

In the case of English, the specific research field of vowels must be sometimes further broken down for analytical purposes because English has a large number of vowel contrasts. Standard British English (SBE) is described to have 12 monophthongal and eight diphthongal vowel phonemes. (This arrangement should not be taken as a strict law. Due to the constant development of sound patterns in English, 13 monophthongs and seven diphthongs are suggested by more recent accounts and different numbers can be found in the description of various English accents, e.g., Ball, 1984; Hawkins & Midgley, 2005.) The reason why foreign accents matter stems from their controversial position in our lives. Although our typical explicit declarations speak about tolerance or indifference, the implicit impact of foreign-accented

speech on perception (and also self-perception) of speakers is indicated by many research results (e.g., Brennan & Brennan, 1981; Bresnahan et al., 2002; Gluszek & Dovidio, 2010; Kavas & Kavas, 2008; Lev-Ari & Keysar, 2010; Lindemann, 2003; Miller & Hewgill, 1964; Rubin, 1992). It is highly irresponsible to claim that the effects of foreign-accented speech will disappear if we disregard them (see Volín, 2018).

Czech learners of English display various degrees of aptitude and they learn to speak the language with varying success. It could be said that they are familiar with the front mid /e/ and back mid /ɔ:/ English sounds because these two vowels have their close equivalents in the Czech language. In contrast, the open to open-mid /æ/ and /ɒ/ are unusual sounds that they must acquire if they wish to speak without a conspicuous foreign accent. Given that the sound contrasts of /e/ vs. /æ/ and /ɔ:/ vs. /ɒ/ are seemingly analogical, we set out to investigate to what extent these contrasts contributed to the perception of Czech-accented English. Inspired by a study (Torstensson et al., 2004) that asked native Swedish speakers to mimic Swedish spoken by native British English speakers, we asked a group of Czech learners of English to produce the English they considered their best self-representation and to mimic a "typical" Czech accent of English. Our premise was that Czech learners could share a view or cognitive prototype of the properties of the English-accented Czech just as the Swedes in that study did with English-accented Swedish. To this aim, we developed an experiment that assessed holistically Czech learners of English accent via external evaluators (e.g., experienced teachers of English) and internal mechanisms (i.e., the learners' judgement of best English vs. Czech accent of English). We then explored the two vowel contrasts in each condition to answer the following research question:

1. To what extent are /e/ vs. /æ/ and /ɔ:/ vs. /ɒ/ likely to participate in external/internal concepts of Czech accent of English?

   a. Do Czech learners of English change their vowel contrasts when asked to enhance their Czech accent of English?
   b. Do three externally assessed subgroups differentiate between the vowel phonemes in opposition?

Another research question that is more distant yet related to our current research is that of symmetricity in phonological inventories. Descriptions of segmental systems of languages mention apparent parallelisms, for instance, between front and back, or open and close vowels. However, distributional and combinatory properties of individual vowel phonemes seldom show any mirroring. Hence, we also sought to investigate any potential analogies between the /e/ × /æ/ and /ɔ:/ × /ɒ/ oppositions in our research context. Specifically, our second research question was:

2. Is there symmetricity in the students' performance, that is, do individual students treat the front and the back contrast analogically in their speech production?

The answer to this question should contribute to the solution of a larger problem concerning competing models of speech representation by human cognitive mechanisms. Symmetrical behaviour would favour deterministic models, in which the

properties of the inventory guide the internal structures. Asymmetricity would favour episodic (or stochastic) models, in which the internal structures are guided by the properties of the actual usage, that is, frequencies of occurrences, repeated contextual dependencies, etc. (cf. Moore, 2007; Pisoni, 1997).

2 Method

2.1 Participants

Thirty-eight university students of various philological programmes (all female, aged 19–24) took part in the experiment. Recording their speech was a routine component of their seminar work for which they were awarded credits. They were all native speakers of Czech, and their competence in English was at the level of B2 to C1 of the CEFR. They were also fluent readers without any hearing problems or speech impairments. They were told that their recordings would be stored anonymously and later analysed, but they were not informed about the exact purpose of the experiment.

During the experiment, participants were assigned to either one of two conditions: internal ($N = 20$) and external ($N = 18$). The former was labelled "internal" (INT) because participants were asked to read the same text twice guided internally to produce less, and later more accented speech. The latter was labelled "external" (EXT) because participants' accentedness was assessed by experienced pronunciation teachers (Skarnitzl et al., 2005), who were not aware of the purpose of the experiment.

2.2 Materials and Procedure

Two news bulletin texts from recent broadcasts in English were used in the experiment. Participants were recorded as they read one of these texts aloud. The recordings took place in a sound-treated studio booth of the Institute of Phonetics at Charles University in Prague. The AKG C4500 B-BC condenser microphone was used connected with the external sound card Steinberg UR44. The signal was captured by Adobe Audition CS6 package at a sampling rate of 32 kHz and an amplitude resolution of 16 bits. Prior to recording, participants were given sufficient time to get acquainted with their texts and were invited to consult any of their possible lexical or grammatical doubts.

The INT group read aloud a shorter text (250 words) twice. They were asked to use their best English pronunciation during the first reading and imitate a typical Czech accent of English during the second reading. Their two renditions were separated by the reading of a Czech text of a comparable length to facilitate the transition to

Czech-accented English. The first reading was labelled "reduced Czech accent of English" (INT-R) and the second "enhanced Czech accent of English" (INT-E).

The EXT group was asked to read aloud a longer text (500 words; seven paragraphs) fluently using their best English pronunciation. Nine experienced teachers of English were asked to rate participants' recordings based on a general impression of their accentedness, where 1 = *Little or No Accent*, 2 = *Ambiguous Accent* and 3 = *Strong Czech Accent*. The raters were not aware of the purpose of this experiment and were explicitly asked not to focus solely on a specific pronunciation feature but rather capture their general feeling. From the original number of 60 recordings we selected 18 with the highest agreement among raters and assigned them to one of three experimental subgroups: "inconspicuous" (EXT-1) ($n = 6$), "moderate" (EXT-2) ($n = 6$), and "strong" (EXT-3) ($n = 6$) Czech accent of English.

2.3   Analyses

After the recordings, individual instances of the four vowels under study—i.e., /e, æ, ɒ, ɔ:/—were identified in the two texts phonemically with respect to SBE, i.e., regardless of the actual sound. This means that the vowel in the word *act* was categorised as /æ/ even if the speaker pronounced clear [ekt]. Similarly, the vowel in *hot* was categorised as /ɒ/ even if it sounded more like [o] or [ɑ]. This is because we were primarily interested in the realizations of the underlying (canonical) segments. However, the auditory impression was not ignored entirely (see Sects. 3.3 and 3.4). The vowel boundaries were manually positioned based on visual and auditory inspection. Only stressed vowels were measured and only those that were not aimed at a different vowel (e.g., items where *acrid* was incorrectly realized as [eɪkrɪd] or *court* as [kɜːt] were excluded).

Formant values were obtained in Praat (Boersma & Weenink, 2016) using the default robust extraction settings for female speakers (10 poles, 5500 Hz upper limit). In total, 480 vowels were analyzed in the INT group and 2145 in the EXT group. A mean value for each formant was calculated from five equidistant values extracted in the central third of each vowel. This procedure reduced the noise in the data caused by transient portions of the vocalic formants at the boundaries of vowels with consonants.

As the formant values correlate to a considerable extent with the position of the tongue in the oral cavity, we refer to vowels with higher values of $F1$ as more open and vowels with higher values of $F2$ as more front. This metonymy is common in current presentation practice, even if the two-dimensional image of the tongue does not explain the vowel acoustics in its entirety. Naturally, to make the correspondence between the sagittal cross-sections of the oral cavity and acoustic plots clear, $F2$ must be placed on the abscissa (x-axis) and $F1$ on the ordinate (y-axis), and the zero value needs to be conceptualized on the top right.

Statistical analyses of the data involved *t*-tests for repeated measures and linear mixed-effects models using the *lme4* package in R followed by calculations of Tukey post-hoc contrasts computed with the *multcomp* package in R (Bates et al., 2015; Hothorn et al., 2008; R Core Team, 2019).

3 Results

This section first presents the results for the INT and EXT groups separately. It then compares the mean formant values obtained in each group to those from a group of native speakers of English reported by Deterding (1997).

3.1 The Internal Assessment Group

The mean formant values captured in Fig. 1 illustrate the situation in the INT sample. The first obvious thing to notice is the smaller distance between the oppositions under the enhanced accent (INT-E) condition: the black targets are closer to each other both for the front pair [e, æ] and for the back pair [ɔ:, ɒ]. In the case of the front vowels, the difference is caused by $F1$ while $F2$ is virtually identical for the white-black pairs. Moreover, $F1$ only matters for the open front [æ]. Generally, when Czech speakers
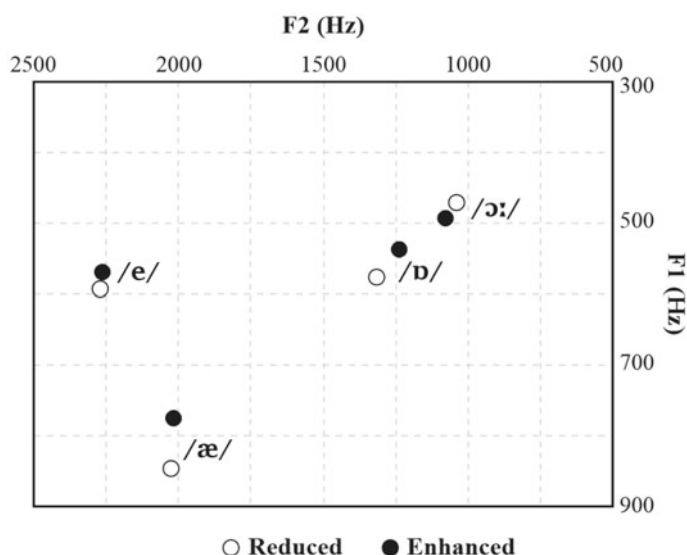

Fig. 1 The vocalic space with vowels /æ, e, ɒ, ɔ:/ produced under the INT-R (reduced Czech accent) and INT-E (enhanced Czech accent) conditions

of English produce [e], the outcome is almost identical under both conditions (*t*-test for repeated measures established the differences for both formants as statistically insignificant). However, under the INT-R condition, [æ] is clearly more open: t (19) = 2.59; $p = 0.018$.

The back opposition does not display any substantial analogy, even though we might say that the mid [ɔ:] is again almost the same under both conditions (the differences for both formants are statistically insignificant), while the open-mid [ɒ] is more open under the INT-R condition. Interestingly though, *t*-tests for repeated measures found significance only for $F2$: t (19) = 2.89; $p = 0.009$. The result for $F1$ was not significant: t (19) = 1.56; $p = 0.134$.

For the purpose of generalization, however, it might be useful to check how individual speakers contribute to this sample outcome. Figure 2 displays the difference in the openness of [æ] between the INT-R and INT-E conditions. It can be observed that only 13 of the 20 speakers produced their [æ] more open when speaking with the reduced accent (INT-R condition), that is, the white column is longer than the black column in the graph. Seven others produced the opposite trend. However, if we turn to perceptual rather than statistical significance, we can evaluate the situation somewhat differently. According to past research in just noticeable differences, the formant values are perceived as different if they are at least five percent from each other (see, e.g., Pols, 1999). Of the 20 speakers, 11 produced noticeably more open [æ], five did not differentiate much, and four produced a closer [æ] under the INT-R condition. In other words, slightly more than a half of the sample indicated the knowledge (whether implicit or explicit) of the required openness of [æ] in SBE pronunciation.

Although the analogical treatment of openness in [ɒ] did not reach statistical significance (as previously shown), we examined it as we did with [æ]. This decision
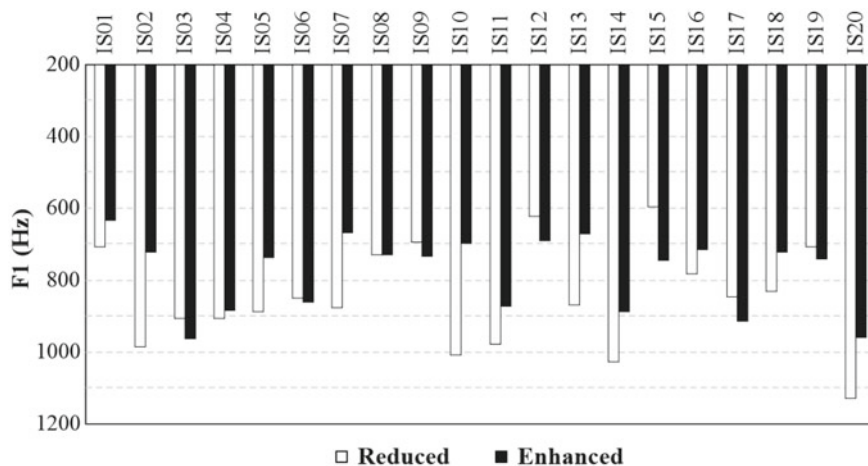

Fig. 2 The mean values of $F1$ for the vowel [æ] by individual speakers under the INT-R (reduced Czech accent) and INT-E (enhanced Czech accent) conditions

Fig. 3 The mean values of *F*1 for the vowel [ɒ] by individual speakers under the INT-R (reduced Czech accent) and INT-E (enhanced Czech accent) conditions
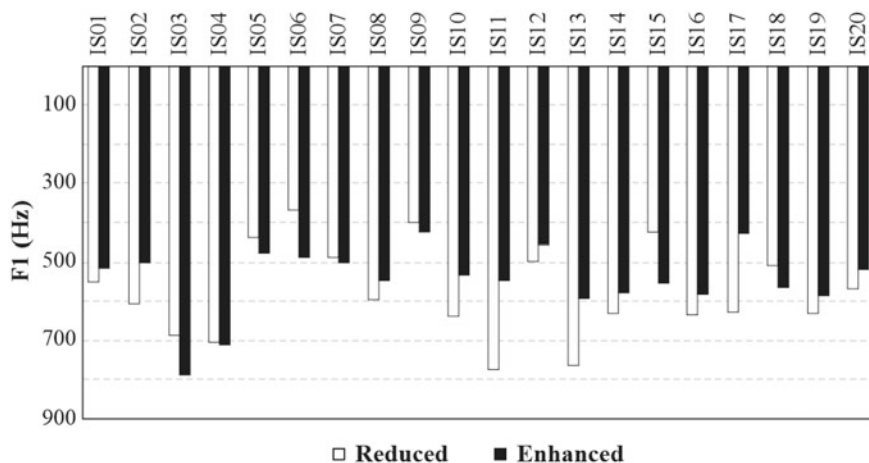
was motivated by the apparent phonological parallelism of /æ/ and /ɒ/ in the system of SBE vowels. Figure 3 displays the values. Similarly to the situation with [æ], 12 speakers produced more open [ɒ] under the reduced accent condition (INT-R), which is the desirable outcome. Nine of those also produced more open [æ] under INT-R (but one did not reach perceptual significance). In abstract terms, these speakers exploited the feature of openness for both the front and the back opposition in order to reduce their accents. Conversely, it is worth noticing that speaker IS15 produced the vowels consistently against the arrangement in the SBE system.

3.2   The External Assessment Group

The EXT speakers, who were externally assessed and divided into three subgroups according to their overall pronunciation competence, produced formant values that resonate to some extent with the situation in the INT group, but add some more information on Czech-accented English.

Figure 4 shows that the difference among the three EXT subgroups is smaller for [e] and [ɔ:] than for [æ] and [ɒ]. A plausible explanation for this is that the vowels [e] and [ɔ:] have their close counterparts in Czech and, therefore, Czech speakers of English are not compelled to modify them in any significant manner. The other two vowels ([æ] and [ɒ]) are more scattered in our vocalic plot, but more importantly, the phonological contrast between neighbours is emphasised in the speech of the best pronouncing subgroup (EXT-1) (white shapes) and curtailed by the worst pronouncing subgroup (EXT-3) (black shapes). The acoustic distance between [e] and [æ], and between [ɔ:] and [ɒ] is clearly the largest for the white shapes in
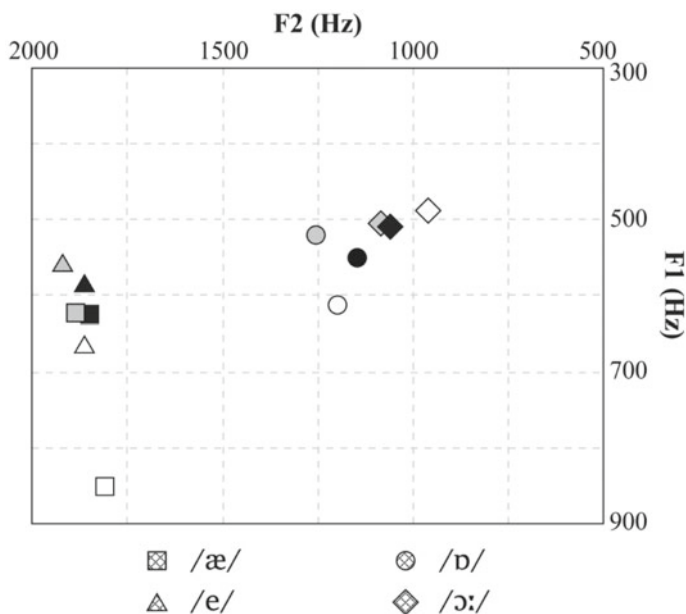
Fig. 4 The vocalic space with vowels /æ, e, ɒ, ɔ:/ produced by EXT-1 (white), EXT-2 (grey) and EXT-3 (black) groups; [e]—triangles, [æ]—squares, [ɒ]—circles, [ɔ:]—diamonds

the graph, that is, for the speakers who achieved the highest pronunciation scores (EXT-1). The 'black group' (EXT-3) produced quite minimal differences, and the 'grey group' (EXT-2) was only marginally better. It should perhaps be repeated at this point that the pronunciation scores were based on the overall impression, that is, they were not focused on vowels.

Since there were enough data points obtained from the larger texts read by the three EXT subgroups, linear mixed-effects models were built using the *lme4* package in R. The individual speakers were set as a random effect (SPEAKER), while PHONEME (i.e., /e, æ, ɒ, ɔ:/) and LEVEL (i.e., EXT-1, EXT-2, and EXT-3) were treated as fixed effects. The *p* values in Table 1 are based on calculations of Tukey post-hoc contrasts (computed with the *multcomp* package in R).

Interaction between LEVEL and PHONEME was established as significant for both $F1$, $\chi^2(6) = 161.2$, $p < 0.001$, and $F2$, $\chi^2(6) = 31.6$, $p < 0.001$. Table 1 summarizes the relevant comparisons between open-mid and mid vowels at the front or back. Both contrasts were significantly different in $F1$ and $F2$ in the EXT-1 subgroup (inconspicuous accent). In the EXT-2 subgroup (moderate accent), the front vowels differed significantly only in $F1$ and back vowels in $F2$. In the EXT-3 subgroup (strong accent), $F2$ did not show any significant differences, and only the front vowels significantly differed in $F1$. However, when taking into account effect sizes, any significant effects in the EXT-2 and EXT-3 subgroups were considerably lower than in the EXT-1 subgroup. Also, openness ($F1$) tended to be associated with greater changes than frontness ($F2$).

Table 1 Pairwise differences between formants of open-mid and mid vowels in EXT subgroups. Adjusted *p*-values from Tukey contrasts in a multiple comparison of means applied to an LME model

Formant	Level	Opposition	Estimate [Hz]	*SE*	*z*	*p*
*F*1	EXT-1	/æ/ – /e/	185	13.5	13.74	< 0.001***
*F*1	EXT-1	/ɒ/ – /ɔ/	101	18.3	5.50	< 0.001***
*F*1	EXT-2	/æ/ – /e/	75	12.4	6.09	< 0.001***
*F*1	EXT-2	/ɒ/ – /ɔ/	−5	17.4	−0.29	1.00
*F*1	EXT-3	/æ/ – /e/	64	12.7	5.01	< 0.001***
*F*1	EXT-3	/ɒ/ – /ɔ/	11	17.1	0.65	0.99
*F*2	EXT-1	/æ/ – /e/	−83	21.9	−3.77	< 0.01**
*F*2	EXT-1	/ɒ/ – /ɔ/	176	29.9	5.89	< 0.001***
*F*2	EXT-2	/æ/ – /e/	−37	19.8	−1.86	0.70
*F*2	EXT-2	/ɒ/ – /ɔ/	161	27.7	5.79	< 0.001***
*F*2	EXT-3	/æ/ – /e/	−51	20.5	−2.46	0.28
*F*2	EXT-3	/ɒ/ – /ɔ/	49	27.3	1.78	0.76

EXT-1 = inconspicuous accent; EXT-2 = moderate accent; EXT-3 = strong accent
* = marginally significant (unaccounted), ** = significant, *** = highly significant

3.3   Two Sources of [æ] Values

When listening to the recordings, we noticed a potential methodological problem. Many of the EXT-2 (moderate accent) and EXT-3 (strong accent) speakers represented the phoneme /æ/ in two disparate ways. In the original Anglo-Saxon words, the common pronunciation was perceptually similar or identical to [e], but in certain internationally used words (often of Latin or Greek origin, but which also have a cognate or near-cognate in Czech), the substituting segment sounded quite like [ʌ] or [a]. Typical examples of the former set are words like *back*, *bank*, *crash*, *man*, *sack*, *tank*, while the latter set can be exemplified by *Africa*, *album*, *Castro*, *collapse*, *commander*, etc. Averaging vowels from both sets could bias the results: we can easily imagine that calculating an arithmetic mean from [e] and [a] will lead to values close to [æ]. Figure 5 shows what happens if the representations of the phoneme /æ/ are measured as three different subsets instead of just one category. The reader is invited to imagine where the average value might lie.

Two important circumstances must be stressed. First, the number of cases in individual categories were largely unbalanced: there were 218 instances of the type /æ/ → [e], but only 76 instances of /æ/ → [ʌ]. Second, the three groups of speakers contributed to the individual categories in a different proportion: the type /æ/→ [æ] was almost exclusively satiated by the speakers with inconspicuous foreign accent (EXT-1) with some small contribution from the speakers with moderate accent (EXT-2).
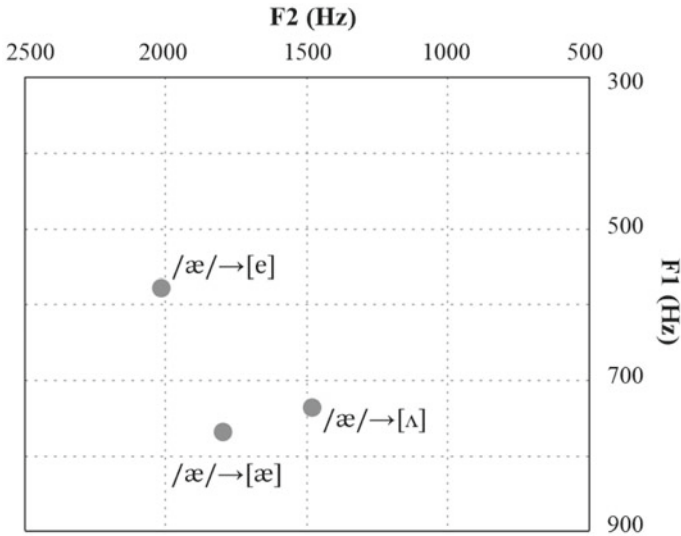
Fig. 5 The vocalic space with three perceptually distinct representations of the phoneme /æ/ found in the vowels produced by the EXT group

Table 2 Standard deviations (in Hz) from mean formant values in vowels corresponding to /e, æ/ produced by INT and EXT speakers

	F1 of /e/	*F1* of /æ/	*F2* of /e/	*F2* of /æ/
INT-speakers	92.3	121.8	126.9	148.0
EXT-speakers	82.3	124.6	126.6	231.4

INT = internally assessed; EXT = externally assessed

The inspection of standard deviation values as a correlate of variability in formant values also suggests that what we identified as [æ] because it represented the phoneme/æ/ is different from [e]. Table 2 indicates that the standard deviations are substantially higher for realizations of /æ/ than of /e/ in both groups (INT and EXT) and both vocalic formants (*F*1 and *F*2).

However, apart from two or three disparate representations of /æ/, the higher variance could also be the result of a less firm articulatory target for this foreign element. It could be argued that the target for /e/ is firmer as it is identical with a frequently occurring vowel in the speakers' L1.

Table 3 Mean values of the first two formants of vowels [e, æ, ɔ:, ɒ] under various conditions

	[e]		[æ]		[ɔ:]		[ɒ]	
	F1	F2	F1	F2	F1	F2	F1	F2
DET	719	2063	1018	1799	389	888	751	1215
INT-R	593	2269	848	2019	472	1041	575	1317
INT-E	571	2258	776	2009	491	1077	538	1234
EXT-1	664	1861	851	1805	492	967	613	1198
EXT-2	557	1920	625	1885	505	1085	519	1250
EXT-3	584	1964	623	1846	509	1069	547	1149

DET = Deterding (1997); INT-R = internally assessed, reduced accent; INT-E = internally assessed, enhanced accent; EXT-1 = externally assessed, inconspicuous accent; EXT-2 = externally assessed, moderate accent; EXT-3 = externally assessed, strong accent

3.4   Comparison with Native English

Although our primary concern in the current study is the relationship of accentedness assessment to vocalic formants in Czech speakers of English, a comparison of the obtained values with the reference values of native speakers of English found in the literature might be of interest. We took our reference values from Deterding (1997), Table 2, columns for female speakers (p. 49). Table 3 displays the summary of the mean values obtained from three groups of Czech speakers (INT-R and INT-E are actually one group under two conditions).

When we compare Deterding's (1997) reference values with our other values within each column, we can observe that in six out of eight cases the closest values to the native speaker sample are those produced by EXT-1, that is, the externally assessed speakers with inconspicuous accent. In two columns only, the closest match comes from the INT-R group: F2 for [e], and F1 for [ɔ:]. This result also corroborates the suggestion that the investigated vowels do play a certain role in impressionistic evaluations of accentedness.

4   Discussion

It is well known that vowels are somehow less phonologically charged than consonants throughout the languages of the world. They withstand greater distortions and their neutralization does not necessarily lead to a collapse in spoken interaction. Therefore, we wanted to investigate whether vocalic contrasts enter the concept of foreign accent held by external assessors or the speakers themselves, even if they are not consciously highlighted.

Our results indicate that under the internal assessment condition, our participants (Czech speakers of English) produced mid vowels [e, ɔ:] virtually the same way for both the reduced and enhanced accent, while their lower vowels [æ, ɒ] differed

in that [æ] was more open, while [ɒ] was more centralized for the reduced accent. This tendency is in line with the requirements of standard pronunciation (SBE). However, individual speakers contributed to this outcome in a non-uniform manner. About one half of the speakers seemed to use the openness feature consistently and quite saliently to reduce their Czech accent, but the rest was apparently either not aware or not in control of the feature. Among the twenty speakers in the INT sample, there was also one striking exception. This speaker (IS15) used openness of [æ] and [ɒ] saliently against the reference SBE system. This result resonates with the findings of Neuhauser (2008), who concluded that her speakers were generally unable to perform consistently, and of Hao and de Jong (2016), who acknowledged that artificial production of an accent was not easy. Conclusions like these, however, should always take into account the speakers' exposure to different accents (cf. Szpyra-Kozłowska, 2018). Our sample comprised speakers who were more than familiar with the Czech accent of English—it was the most frequently heard foreign accent in their experience. Therefore, the individual variation should be attributed to personal sensitivity or acuity of the speakers.

The externally assessed (EXT) group provided data that can be reconciled with the results discussed in the previous paragraph. The subgroup with an inconspicuous Czech accent in their English (EXT-1) produced both vowel contrasts in line with descriptions of the Southern British Standard. Their open-mid vowels were, indeed, more open than their mid vowels. The subgroup with strongly accented English (EXT-3), on the other hand, made very small distinctions between the representations of /e/, /æ/ and /ɔː/, /ɒ/, respectively.

A methodological note has to be made about averaging values for the open front vowel [æ] across the board. Our sample indicated that this could lead to serious flaws since the texts included words of Latin, Greek or other non-Anglo-Saxon origin where many Czech speakers do not even attempt at [æ]. In our analysis, we observed that the /æ/ was pronounced as [e] in English words with no equivalent in Czech, while the /æ/ was pronounced as [ʌ] or [a] (central open-mid or open vowel) in some 'international' words. This is a relatively new problem in our line of research. In the past, vowel formants were often measured in restricted sets of words that were recorded in isolation, and not in open sets of continuous meaningful texts.

A similar problem is posed by the consonantal influence on vowels. Rather than carefully pronounced sets like *heed – hid – head – had*, etc., continuous texts comprise various combinations of consonants with vowels, and their effects are not necessarily cancelled out because the phonotactic properties of vowels are not balanced in the phonological system of the English language. Specifically in our case, the back mid vowel /ɔː/ was very often followed by the liquids /r, l/, both of which are known to pull formants in neighbouring vowels down (e.g., Volín & Studenovský, 2007). Therefore, the physical contrast we found between the pronounced [ɒ] and [ɔː] in our sample could have been exaggerated by the liquid effects. However, this bias is part of the sound structure of English, so trying to neutralize it might make certain results less ecologically valid.

It follows that when making generalizations about foreign-accented speech, the type of speech material has to be taken quite seriously (cf. also Thompson, 1991, for a similar conclusion). Another caveat to be considered concerns the representativeness of our sample. First of all, it has to be stressed that the sample consisted of philologically oriented university students with positive attitudes to English. Therefore, it does not represent the Czech population as such. Moreover, the EXT sample was artificially balanced to comprise equal numbers of speakers in each subgroup. It can be hardly expected, however, that one third of the Czech population of learners of English is formed by speakers with an inconspicuous accent. It is quite realistic to expect that people with a strong accent prevail. Nevertheless, these considerations are not crucial for our research question. We set out to find out whether there is a link between vowel contrasts as expressed in formant measurements and criteria of general accentedness in Czech English. Our research indicates that such a link exists.

5 Implications

The results of our experiment contribute to two major areas of expertise: (a) didactics and design of classroom practice, and (b) theoretical models of speech processing. As to the first one, although the vowel qualities seem to correlate with both the internal and external evaluations of foreign accented speech, a substantial number of learners was apparently not aware of the difference. Even those who were, did not produce the contrasts strictly in parallel. Generally speaking, learners of English would most probably benefit from practicing various pairs of contrasts interlinked by a shared distinctive feature. In our case this would mean that rather than training the pronunciation of [e] vs. [æ] on its own, and sometime later focusing on the isolated pair [ɔ:] vs. [ɒ], the students should practice both contrasts together because they are both based on the feature of openness. In other words, pronunciation training should not be based on isolated phonemic contrasts but on pairs or groups of contrasts that share a salient feature.

As to the second area, the lack of strict symmetricity in our data seems to support the foundation of the episodic models of speech perception. These models maintain that rather than the abstract properties of the phonological inventory, human speech processing relies on the stochastic features of language use. If individual items of the inventory occur in different phonotactic combinations, situational contexts or types of lexicon, they will be treated differently despite the apparent analogy in the inventory arrangement. In classrooms, this would imply advantage of generous language input (namely active listening) over explanations of language structures.

6 Conclusions

We hope that our investigation, which provided formant measurements for cross linguistic comparisons, has contributed to the emerging picture of the mechanisms of foreign accents in speech. We believe that it would be beneficial to the understanding of second language acquisition to contrast the performance of Czech students of English to that of speakers of other languages. However, to draw any useful analogies, our methodology has to be replicated as closely as possible. We especially warn against using material based on pronunciation of isolated words or contextless sentences.

Our study focused on two apparently parallel phonological contrasts: /e/ vs. /æ/ and /ɔ/ vs. /ɒ/. We found out that their resonance with holistic assessments of accentedness is stronger in external conditions (in our case by experienced teachers). Internal assessments (by the speakers themselves) involved the investigated vowel contrasts in a smaller scale. It follows that if the L2 learning objectives include acquisition of sound patterns of the target language, the learners must be directed to the specific vocalic contrasts through the teaching process. Our data suggest that the problem is resolved spontaneously in only about one third to one half of the population. Other learners need external help.

Naturally, there is always the possibility to resign on pronunciation training and leave the accentedness unaltered. Unfortunately, the consequences of such a decision are still poorly understood since the research in this area is often biased by wishful thinking. What is ultimately needed are rigorously planned perception experiments informed by production studies like ours.

References

Ball, M. (1984). The centring diphthongs in Southern English: A sound change in progress. *Journal of the International Phonetic Association, 14*(1), 38–44. https://doi.org/10.1017/S00251003000 02723

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beňuš, Š. (2012). Phonetic variation in Slovak yer and non-yer vowels. *Journal of Phonetics, 40*(3), 535–549. https://doi.org/10.1016/j.wocn.2012.03.001

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Version 6.0.19) [Computer software]. http://www.praat.org

Brennan, E. M., & Brennan, J. S. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research, 10*, 487–501. https://doi.org/10.1007/BF01076735

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language and Communication, 22*(2), 171–185. https://doi.org/10.1016/S0271-5309(01)00025-8

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association, 27*(1–2), 47–55. https://doi.org/10.1017/S0025100300005417

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America, 126*(3), 1379–1393. https://doi.org/10.1121/1.3180321

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association, 40*(1), 1–34. https://doi.org/10.1017/S0025100309990247

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, *126*(5), 2603–2618. https://doi.org/10.1121/1.3212921

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review, 14*(2), 214–237. https://doi.org/10.1177/1088868309359288

Hao, Y. C., & de Jong, K. (2016). Imitation of second language sound in relation to L2 perception and production. *Journal of Phonetics, 54*, 151–168. https://doi.org/10.1016/j.wocn.2015.10.003

Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association, 35*(2), 184–199. https://doi.org/10.1017/S0025100305002124

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099–3111. https://doi.org/10.1121/1.411872

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Kavas, A., & Kavas, A. (2008). An exploratory study of undergraduate college students' perceptions and attitudes toward foreign accented faculty. *College Student Journal, 42*, 879–890.

Leppik, K., Lippus, P., & Asu, E. L. (2019). The production of Estonian vowels in three quantity degrees by Spanish L1 speakers. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren *Proceedings of 19th International Congress of Phonetic Sciences* (pp. 1154–1158). IPA.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about nonnative English speakers in the United States. *Journal of Sociolinguistics, 7*(3), 348–364. https://doi.org/10.1111/1467-9481.00228

Miller, G. R., & Hewgill, M. A. (1964). The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech, 50*(1), 36–44. https://doi.org/10.1080/00335636409382644

Moore, R. K. (2007). Spoken language processing: Piecing together the puzzle. *Speech Communication, 49*(5), 418–435. https://doi.org/10.1016/j.specom.2007.01.011

Munro, M. J., Derwing, T. M., & Saito, K. (2013). English L2 vowel acquisition over seven years. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 112–119). Iowa State University.

Neuhauser, S. (2008). Voice disguise using a foreign accent: Phonetic and linguistic variation. *The International Journal of Speech, Language and the Law, 15*(2), 131–159. https://doi.org/10.1558/ijsll.v15i2.131

Pisoni, D. B. (1997). Some thoughts on 'normalization' in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). Academic Press.

Pols, L. C. W. (1999). Flexible, robust, and efficient human speech processing versus present-day speech technology. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of 14th International Congress of Phonetic Sciences* (pp. 9–16). UCLA. https://www.fon.hum.uva.nl/archive/1999/icphs/PolsICPhS1999.pdf

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 3.5.3). R Foundation for Statistical Computing, Vienna. Retrieved August 26, 2020, from www.r-project.org

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*, 511–531. https://doi.org/10.1007/BF00973770

Šimáčková, Š, & Podlipský, V. J. (2018). Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility. *Research in Language, 16*(2), 169–191. https://doi.org/10.2478/rela-2018-0009

Skarnitzl, R., & Volín, J. (2012). Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické Listy, 18*(1), 7–11.

Skarnitzl, R., Volín, J., & Drenková, L. (2005). Tangibility of foreign accents in speech: the case of Czech English. In A. Grmelová, L. Dušková, & M. Farrell (Eds.), *Proceedings of the 2nd Prague Conference on Linguistics and Literary Studies* (pp. 11–20). UK PedF.

Šturm, P., & Skarnitzl, R. (2011). The open front vowel /æ/ in the production and perception of Czech students of English. In P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini (Eds.), *Proceedings of 12th Interspeech 2011* (pp. 1161–1164). ISCA. https://doi.org/10.21437/Interspeech.2011-344

Szpyra-Kozłowska, J. (2018). The rhotic in fake and authentic Polish-accented English. *Lublin Studies in Modern Languages and Literature*, *42*(1), 81–102. https://doi.org/10.17951/lsmll.2018.42.1.81

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning, 41*(2), 177–204. https://doi.org/10.1111/j.1467-1770.1991.tb00683.x

Torstensson, N., Eriksson, E. J., & Sullivan, K. P. H. (2004). Mimicked accents. Do speakers have similar cognitive prototypes? In S. Cassidy, F. Cox, R. Mannell, & S. Palethorpe (Eds.), *Proceedings of the 10th Australian International Conference on Speech Science & Technology* (pp. 271–276). ASSTA.

Volín, J. (2018). Foreign accents and responsible research. In J. Volín & R. Skarnitzl (Eds.), *The pronunciation of English by speakers of other languages* (pp. 4–18). Cambridge Scholars Publishing.

Volín, J. & Studenovský, D. (2007). Normalization of Czech vowels from continuous read texts. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 185–190). IPA a UDS. http://www.icphs2007.de/conference/Papers/1722/1722.pdf

Jan Volín is Associate Professor at the Institute of Phonetics in Prague, Czech Republic. His research focuses on sound structure of languages, general issues of speech prosody, speech acquisition, and on various pragmatic and psychological aspects of human speech behaviour. These topics alternate in his papers published in various scientific journals, chapters in monographs, and handbooks dedicated to suprasegmental phonetics.

Tanja Kocjančič Antolík's research is focused mainly on speech production in typical and atypical speech. She has been involved in different research projects across Europe addressing speech in a foreign language learning and speech disorders. She employs acoustic and articulatory analysis and is an expert in ultrasound tongue imaging.

Radek Skarnitzl is Associate Professor at the Faculty of Arts, Charles University, Prague, Czech Republic. His research focuses on second language pronunciation and the effect of various

pronunciation features on the socio-psychological evaluation of a speaker in both native and foreign languages. He is also interested in issues related to speaker identification, especially the effects of voice disguise.

Pavel Šturm is Assistant Professor at the Faculty of Arts, Charles University, Prague, Czech Republic. His research interests lie mainly in the domain of second language pronunciation (especially L2 English), focusing on segmental aspects in terms of production and on foreign accent evaluation and perception. He also examines the issues of syllabification and phonotactics in the Czech language.