

SPECTRAL CHARACTERISTICS OF SCHWA IN CZECH ACCENTED ENGLISH

JAN VOLÍN

Metropolitan University Prague
Jan.Volin@ff.cuni.cz

LENKA WEINGARTOVÁ

Institute of Phonetics in Prague
Lenka.Weingartova@ff.cuni.cz

RADEK SKARNITZL

Institute of Phonetics in Prague
Radek.Skarnitzl@ff.cuni.cz

Abstract

The English central mid lax vowel (i.e., schwa) often contributes considerably to the sound differences between native and non-native speech. Many foreign speakers of English fail to reduce certain underlying vowels to schwa, which, on the suprasegmental level of description, affects the perceived rhythm of their speech. However, the problem of capturing quantitatively the differences between native and non-native schwa poses difficulties that, to this day, have been tackled only partially. We offer a technique of measurement in the acoustic domain that has not been probed properly as yet: the distribution of acoustic energy in the vowel spectrum. Our results show that spectral slope features measured in weak vowels discriminate between Czech and British speakers of English quite reliably. Moreover, the measurements of formant bandwidths turned out to be useful for the same task, albeit less direct.

Keywords: foreign accent, prominence, schwa, spectral slope.

1. Introduction

Foreign accents manifest themselves in various domains of speech architecture. In the segmental area, individual vowels and consonants may display interesting deviations from what is perceived as standard or canonical for native speech, while the specific suprasegmental or prosodic features cause no less remarkable differences in stress, intonation or rhythm. The consequences of sounding foreign can be manifold: from moderate amusement on the part of the native listener to outright prejudice on the attitudinal level, or, in the domain of information flow, from increased strain in speech processing to miscomprehension leading to the breakdown of communication.

Derwing and Munro (2009) summarize some of their observations on the social consequences of foreign accented speech in their general overview and point out that it is

the lack of deeper understanding which sustains negative attitudes to foreign accents. Lev-Ari and Keysar (2010) highlight another source of complications for speakers with a foreign accent: their experiment showed that the credibility of a speaker's statements is affected by the degree of foreign accentedness. The assertions in sentences pronounced by foreigners were perceived as less true than the identical propositions spoken by native speakers of English. However, the design of their study enabled the authors to question the exclusive influence of prejudice. Having analyzed their results, they suggested that the decrease of credibility was connected with increased processing load. This effect can be observed even in the visual domain: if some printed statements are more difficult to read than others due to the colour and size of the font, the perceivers tend to assume they are less true (Lev-Ari and Keysar, 2010). Be that as it may, it is clear that not only areas like second language education, but also forensic practice and security or business would profit from a deeper insight into the mechanisms which underlie communication in non-native languages.

Our study focuses on the English lax central mid vowel, also known as *schwa* or reduced vowel. Due to its unique properties, both phonetic and phonological, this vowel is responsible for specific perceptual effects in the sphere of speech rhythm (along with stress and intonation, even if its impact there is less direct). In this sense, the segmental and suprasegmental domains are brought together quite firmly in our study, through schwa as the object of investigation.

We will not enter the dispute over the phonemic status of schwa. It is clear that, on the one hand, this vowel may participate in a phonological opposition with another vowel: *affect* – *effect*: /ə'fekt/ – /i'fekt/; *boxers* – *boxes*: /'bɒksəz/ – /'bɒksɪz/, or with a phonological zero: *data* – *date*: /dɛtə/ – /dɛt/, on the other hand, there are countless examples of positionally conditioned alternation of schwa with other vowels: *academy* – *academic*: /ə'kædəmi/ – /ækə'demɪk/; *land* – *England*: /'lænd/ – /'ɪŋɡlənd/, etc. Depending on the stress-pattern of the word, speakers may produce either schwa or a full vowel in the given position: the vowel between /d/ and /m/ in the word *academy* can be described as reduced, while the same position in the word *academic* is occupied by the peripheral front mid vowel.

This type of positional conditioning, however, manifests itself not only in derivational processes on the word level, but also on the phrase level. The metrical structure of syntactic units requires a certain arrangement of strong and weak positions. There are about forty monosyllabic grammatical words in English which regularly occur in unstressed positions and, consequently, surface in the so-called weak forms, which in most cases means with schwa as their syllable nucleus. Such words served as the material in our study (see section Method) since despite their small number in the lexicon they are very frequent in texts. They include prepositions (*at, for, from, of, to*), conjunctions (*and, but, than, as*), pronouns (*you, your, her, them, etc.*), auxiliary and modal verbs (*are, were, was, have, has, do, does, would, can, etc.*), and determiners (*a, the, some*). It has to be emphasised that these words sometimes do occur in their strong forms (e.g., when syntactically stranded or under focus) and when they do, full or peripheral vowels are employed in their pronunciation.

Due to the objective of our present study, we can leave open the question of whether schwa is an independent phoneme or an allophone of other vowels. As Ashby *et al.* (1995) pointed out, to most practically minded people the important fact is that schwa is an essential component of the sound patterning of English. According to Fry, almost one

quarter of all vowels in continuous speech correspond to schwa. Just for comparison, if a learner of English does not pronounce the front open vowel /æ/ (also known as *ash*) correctly, only one in twenty-seven vowels is affected in running texts (Fry, 1947).

To a laboratory acoustician, the male formant frequencies of schwa are $F1 = 500$ Hz, $F2 = 1,500$ Hz, and $F3 = 2,500$ Hz. The ideal female values are about fifteen percent higher. However, the evidence collected over the years of phonetic research confirms that the formant values of schwa retrieved from continuous speech are very variable and sensitive to context (e.g., Lindblom, 1963; Browman and Goldstein, 1992; Flemming and Johnson, 2007). Some of these studies showed that the formant positions of schwa are influenced not only by the neighbouring consonants, but also by vowels in the neighbouring syllables. Thus, the listener's feeling of the weak syllable nuclei being properly reduced or not does not seem to be correlated with the ideal position of formants in the "acoustically pure" schwa. Moreover, Barry (1998) put forward some evidence that computational methods used to determine the properties of schwa from formant frequencies were not entirely correct. He also confirmed that the contextual influence on formants is related to articulation rate.

Despite the fact that formant frequencies are often the only spectral descriptors of vowels (or sonorants in general) offered by textbooks on speech acoustics, we have to ask whether there is indeed no other information in the spectrum that could be correlated with the "proper" sound of schwa. A cue that such information should exist is provided by the metrical role of the English schwa. It is a vowel with the inherent feature [+*unstressed*] and in the four-level analysis of syllabic prominences it is associated with the lowest level. It has been pointed out previously that the salience of the vowel is reflected by its spectral slope (also spectral tilt or spectral balance). The brighter sounding vowels display more moderate decay of spectral energies in higher-frequency regions or, the other way round, weaker sounding vowels have steeper decrease of amplitude towards the higher frequencies (Sluijter and van Heuven, 1996; Gobl and Ni Chasaide, 2003).

One of the major problems is that there is no single established method of quantifying spectral slope. One of the early attempts to provide an index of spectral balance was that of Britta Hammarberg and her colleagues who used the difference between the energy peaks (maxima) in the 0–2 kHz and 2–5 kHz frequency bands (Hammarberg *et al.*, 1980: 448). Various modifications were later suggested to this approach. The so-called α measure is based on the ratio between the sound energy above and below 1,000 Hz (e.g., Sundberg and Nordenberg, 2006). In their overview, Hanson *et al.* (2001) discuss several further measures, each with some advantages and disadvantages in the context of various objectives (usually the detection of voice disorders or quantifying long-term average spectra). After experimenting with individual vowels produced by healthy voices, we decided to modify the existing measures (Volín and Zimmermann, 2011), as described below in the Method.

2. Method

Our dataset consisted of recordings of three female native Southern Standard British English speakers and three female native Czech speakers. None of them had a speech

impediment or reported any hearing disorder. Both the native and Czech speakers of English were not professional speakers, and their age ranged from 20 to 23 years. They were asked to read out a single news bulletin taken from a broadcast of the BBC World Service. The three Czech subjects were selected on the basis of two criteria: they had to be capable of reading the relatively difficult English text fluently but, at the same time, they had to exhibit consistently a relatively overt Czech accent.

The British speakers were recorded with a portable professional device Edirol HR-09, with a sampling frequency of 48 kHz and 16-bit resolution. Later, these recordings were resampled to 32 kHz. The recordings of Czech speakers were made in a sound-treated studio of the Institute of Phonetics in Prague with an electret microphone IMG ECM 2000, soundcard SB Audigy 2 ZS, 32-kHz sampling frequency and 16-bit resolution.

This material yielded approximately 4–4.5 minutes of speech for each of the subjects. The utterances were cut into breath-groups and manually labelled in Praat (Boersma and Weenink, 2012) by experienced phoneticians.

The spoken texts contained a total of 1,452 occurrences of schwa, of which 692 were found in monosyllabic grammatical words only (see above). These were selected for further analyses, yielding about 115 schwas per speaker.

The extracted parameters were as follows:

- average F1 and F2: measured in the middle third of each vowel
- average bandwidths of F1 and F2: measured in the middle third of each vowel
- spectral slope: measured as band energy and band density difference in the frequency bands of 350–1,100 Hz and 2,300–5,500 Hz in the middle of each vowel

The extraction of all parameters was done with Praat scripts.

Our method of determining the values of spectral slopes deserves some explanation. Similarly to Hammarberg *et al.* (1980), we calculate the difference in energy between specific frequency bands. In the study of Volín and Zimmermann (2011), a similar method was successfully used to distinguish stressed from unstressed vowels of three Czech speakers. The authors hypothesized that excluding the bands which correspond to F0 and F2 will improve the discriminatory power of the method, since the enormous spectral energy of F0 could cloud the relevant spectral measurements, and the highly variable energies in the F2 region are expected to code mainly vowel identity.

In the current study, the Praat predefined functions of band energy difference and band density difference were used, which calculate the sum (in case of the former) or the average (in case of the latter) energy in the given high-frequency band and subtract it from the low-frequency band. Figure 1 illustrates the measurement.

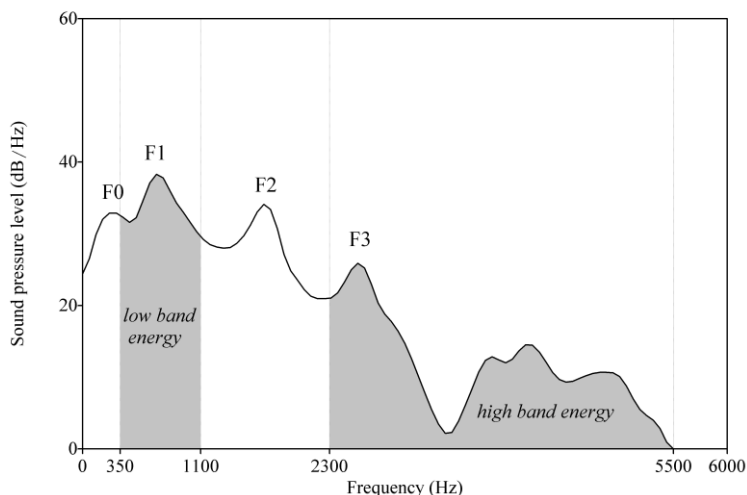


Figure 1. Spectrum of a vowel with highlighted low (350–1,100 Hz) and high (2,300–5,500 Hz) frequency bands. The ranges of F0 and F2 are excluded from the measurement (see text).

To assess the statistical validity of the results, one-way analysis of variance (ANOVA) was employed with the extracted parameters as dependent variables and L1 of the speaker (English/Czech) as a factor.

3. Results

The values of the first two formants did not turn out to be a reliable correlate of the difference between British and Czech speakers. In both cases, the results were non-significant at the level of $p = 0.05$ ($F(1, 690) = 0.0011$; $p = 0.97$ for F1 and $F(1, 690) = 2.62$; $p = 0.11$ for F2).

The bandwidths performed substantially better: both F1 and F2 bandwidths were able to distinguish the British from Czech speakers with high statistical significance (see Figure 2a-b). The ANOVA result for F1 bandwidth was $F(1, 690) = 6.14$; $p = 0.013$ and for F2 bandwidth $F(1, 690) = 12.96$; $p < 0.001$.

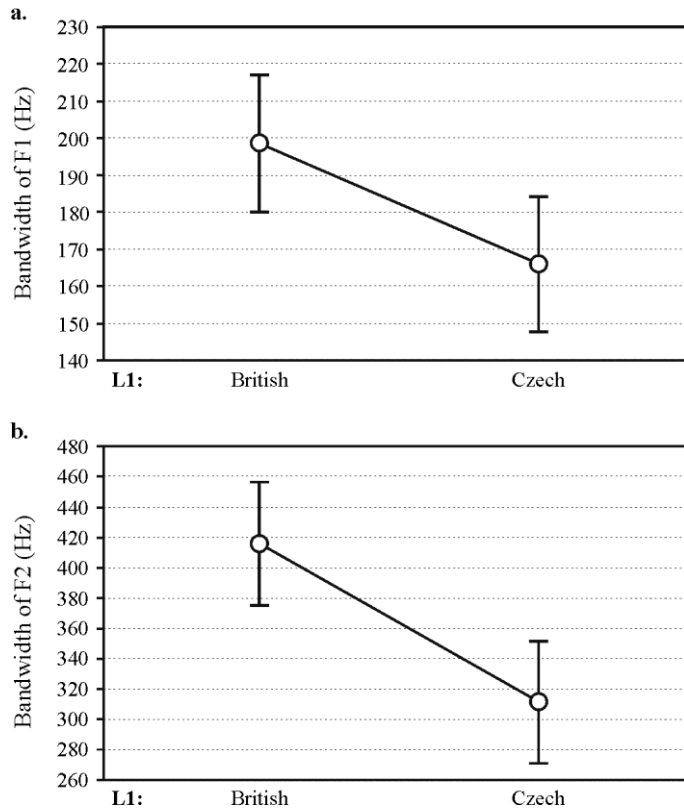


Figure 2. Average values of F1 bandwidth (a.) and F2 bandwidth (b.) of British vs. Czech schwas. Whiskers denote the 0.95 confidence interval.

In both cases, our Czech speakers produced narrower formant bandwidths, which should generally be interpreted as stronger or better defined formants. Nevertheless, these results have to be rationalized with some caution, since the difference in recording conditions of the two groups of speakers (i.e., the frequency response of the recording room) may have slightly influenced the bandwidth values.

The two measures of spectral slope, band energy and band density difference, yielded very similar results and almost identical values of the test criterion F in the inferential analysis of variance. Since density seems to be a better option for computational reasons (it is not sensitive to changes in the frequency band boundaries), we will only present band density difference results for the sake of simplicity. As can be seen in Figure 3, the difference in spectral slope between British and Czech speakers also turned out to be significant ($F(1, 690) = 5.13$; $p = 0.024$).

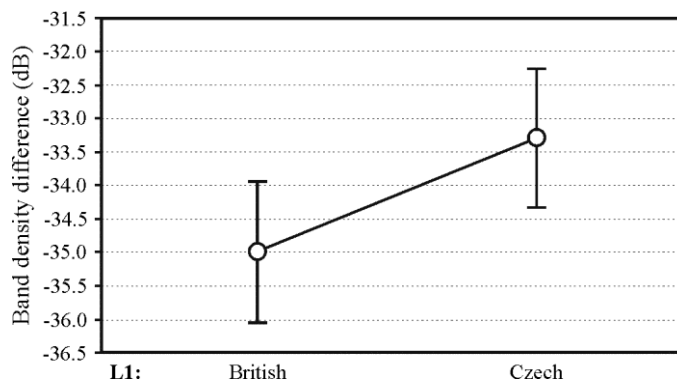


Figure 3. Average values of band density difference (spectral slope measure) of British vs. Czech schwas. Whiskers denote the 0.95 confidence interval.

Figure 3 shows the difference in average values of spectral balance between the British and Czech speakers, where the Czech speakers produced values closer to zero (the scale being negative) of the band density difference, meaning a flatter spectral slope. The spectral slope in schwas of the British speakers was steeper, with less energy in higher frequencies. This is found in less prominent vowel sounds.

4. Discussion

The English central mid lax vowel is relatively frequent in spoken texts. Due to its metrical role (the lowest level of prominence), it is an important element of the rhythmic structure of English. Czech speakers use typologically different rhythm in their mother tongue, and vowel reduction does not take place in the standard pronunciation of Czech. It can be expected that their foreign grasp of English would reflect this situation.

Our study showed that the formant values, despite being the primary descriptors of vowel quality, do not allow for discrimination between native (British) and Czech speakers of English. However, the measurements of spectral slope led to a statistical difference between the average schwa production of British and Czech speakers who were reading out identical texts under identical conditions from the point of view of the communicational context. The Czech schwas displayed a less steep decrease of energy towards the higher frequencies, which means more prominent vowel sounds.

The current results also revealed a significant difference in formant bandwidths. It was especially the second formant that was produced wider by the native speakers than by the non-natives. Since greater bandwidths generally indicate weaker formants, it could be argued that this result is a confirmation of the same phenomenon as the spectral slope measurements. We believe it is. However, it has to be remembered that determining formant bandwidths is computationally much more vulnerable than detecting energy in the spectrum and, also, the formant bandwidths might be more sensitive to recording conditions.

In our future research, we would like to take a closer look at individual analyzed items in the recorded texts. For instance, it could be the case that prepositions are more

useful in discrimination between British and Czech speakers than conjunctions due to the fact that they are more consistent in occurring at weak metrical positions. Some speakers tend to hesitate on conjunctions or fortify them as a signal of a new syntactic unit. Similar, but less obvious relationships might be found for auxiliary verbs and pronouns. In addition, our material also provided over 700 schwas in polysyllabic autosemantic words. These vowels will be analyzed using the same methodology as was used for the current sample after a system of finer classification is developed for the lexical items.

Acknowledgements

This study was supported by the Programme of Scientific Areas Development at Charles University in Prague (PRVOUK), subsection 10 – Linguistics: Social Group Variation. The second author was supported by Internal Grant of the Faculty of Arts, Charles University in Prague (VG192).

References

- Ashby, M., Ashby, P., Baldwin, J., Holmes, F., House, J. and Maidment, J. 1995. Broad transcription in phonetic training. *Proceedings of the XIIIth ICPHS*: 170–173. Stockholm: IPA.
- Barry, W. J. 1998. Time as a factor in the acoustic variation of schwa. *Proceedings of 5th Int. Conf. on Spoken Language Processing*, Sydney: 3071–3074.
- Boersma, P. and Weenink, D. 2012. *Praat: doing phonetics by computer* (version 5.3.14). Retrieved from <http://www.praat.org/>.
- Browman, C. P. and Goldstein, L. 1992. “Targetless” schwa: an articulatory analysis. In G. J. Docherty and D. R. Ladd (eds.) *Papers in Laboratory Phonology II*. Cambridge: CUP: 26–56.
- Derwing, T. M. and Munro, M. J. 2009. Putting accent in its place: rethinking obstacles to communication. *Language Teaching* 42 (4): 476–490. DOI: 10.1017/S026144480800551X
- Flemming, E. and S. Johnson. 2007. Rosa’s roses: reduced vowels in American English. *Journal of the International Phonetic Association* 37: 83–96. DOI: 10.1017/S0025100306002817
- Fry, D. B. 1947. The frequency of occurrence of speech sounds in Southern English. *Archives Néerlandaises de Phonétique Expérimentale* 20: 103–106.
- Gobl, Ch. and Ní Chasaide A. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40: 189–212. DOI: 10.1016/S0167-6393(02)00082-1
- Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J. and Wedin, L. 1980. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngologica* 90: 441–451.
- Hanson, H. M., Stevens, K. N., Kuo, H-K. J, Chen, M.Y. and Slifka, J. 2001. Towards models of phonation. *Journal of Phonetics* 29: 451–480. DOI: 10.1006/jpho.2001.0146

- Lev-Ari, S. and Keysar, B. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 46: 1093–1096. DOI: 10.1016/j.jesp.2010.05.025
- Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 143–162.
- Nakatani, L. H., O'Connor, K. D. and Aston, C. H. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84–105.
- Sluijter, A. M. C. and Van Heuven, V. J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100: 2471–2485.
- Sundberg, J. and Nordenberg, M. 2006. Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech. *Journal of the Acoustical Society of America* 120/1: 453–457. DOI: 10.1121/1.2208451
- Volín, J. and Zimmermann, J. 2011. Spectral slope parameters and detection of word stress. *Technical Computing Prague*: 125 – 5 pages. Praha: Humusoft.