HAL
open science

The strength of foreign accent in Czech English under adverse listening conditions

Jan Volín, Radek Skarnitzl

▶ **To cite this version:**

Jan Volín, Radek Skarnitzl. The strength of foreign accent in Czech English under adverse listening conditions. Speech Communication, 2010, 52 (11-12), pp.1010. 10.1016/j.specom.2010.06.009 . hal-00698872

HAL Id: hal-00698872

https://hal.science/hal-00698872v1

Submitted on 18 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

The strength of foreign accent in Czech English under adverse listening conditions

Jan Volín, Radek Skarnitzl

PII: S0167-6393(10)00117-2
DOI: 10.1016/j.specom.2010.06.009
Reference: SPECOM 1905

To appear in: *Speech Communication*

Received Date: 4 August 2009
Revised Date: 25 June 2010
Accepted Date: 26 June 2010

Please cite this article as: Volín, J., Skarnitzl, R., The strength of foreign accent in Czech English under adverse listening conditions, *Speech Communication* (2010), doi: 10.1016/j.specom.2010.06.009

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

The strength of foreign accent in Czech English under adverse listening conditions

Jan Volín *, Radek Skarnitzl

Institute of Phonetics, Faculty of Arts, Charles University in Prague
nam. Jana Palacha 2, 11638 Praha 1, Czech Republic
jan.volin@ff.cuni.cz, radek.skarnitzl@ff.cuni.cz

* Corresponding author. Tel.: +420 221 619 379; fax: +420 221 619 260.
E-mail address: jan.volin@ff.cuni.cz

Abstract
The study connects two major topics in current speech research: foreign accentedness and speech in adverse conditions. We parallel the research in intelligibility of non-native speech, but instead of linguistic unit recognition we focus on the perception of the foreign accent strength. First, the question of type and degree of perceptual deficiencies occurring along with certain types of signal degradation is tackled. Second, we measure correlations between the accent ratings and certain candidate phenomena that may influence them, e.g., articulation rate, temporal patterning, contrasts in sound pressure levels on selected syllables and F0 variation. The impacts of different types of signal degradation help to estimate the role of segmental/suprasegmental information in assessments of foreignness in Czech English. The full appreciation of the strength of foreign accent is apparently not possible without fine phonetic detail on the segmental level. However, certain suprasegmental features of foreignness are robust enough to manifest at severe levels of signal degradation. Pairwise variability indices of vowel durations and variation in F0 tracks seem to guide the listener even better in the degraded than in the 'clean' speech signal.

Key terms: foreign accent; Czech English; low-pass filtering; perception; rhythm metrics; signal-to-noise ratio.

1. Introduction

A foreign accent is traditionally described as a set of pronunciation patterns, at both segmental and suprasegmental levels, which differ from pronunciation patterns found in the speech of native speakers. Foreign or non-standard features in speech have fascinated people for centuries or perhaps millennia. Apparently, they can even become a source of unscrupulous profit, which builds on the lack of our deeper understanding of their foundation and on the general ignorance of what has already been discovered about them (see, e.g., criticism by Derwing and Munro, 2009). Current research objectives in speech accents are motivated by a number of practical considerations. Some are explained, for instance, in Eskenazi (2009), Ikeno and Hansen (2006), Strik (2003), or Wu et al. (2010). It is clear that areas like second language education, forensic practice and security or business would profit from a deeper insight into the mechanisms which underlie communication in non-native languages.

The earlier emphasis on the pronunciation of foreign segments was expanded to foreign accent perception, which highlighted some very important issues. The most pressing one seems to be the question of false beliefs about the speaker (Pickering, 2001) or even prejudice, which was shown to affect human behaviour startlingly deeply even at subconscious levels (e.g., Rubin, 1992). Further studies mapping the general bias against foreign accents are quoted in Brennan and Brennan (1981), Munro and Derwing (1995: 74),

ACCEPTED MANUSCRIPT

Hahn (2004: 205) and Munro et al. (2006: 112). Thus, whether we face prejudice which impairs interhuman cooperation on the one hand, or claims that the question of accentedness should be completely abandoned as irrelevant or artificial on the other hand, we find the lack of reliable information on the subject matter as the common denominator.

The research of the past decades made a useful link between foreign accent research and speech signal degradation experiments. First, this step is a response to technological demands: automatic speech recognizers seem to profit from correct accent detection (Ghesquiere and Van Compernolle, 2002) and because most of our communication actually takes place in acoustically adverse environments, it is beneficial to know which features suffer least when the speech signal is imperfect. Second, since diverse conditions may affect the non-native users of a language code differently from the native ones, there is an opportunity to pose interesting research questions. For instance, non-native listeners' performance often deteriorates faster with increasing distortion (e.g., García Lecumberri and Cooke, 2006; Mayo *et al*., 1997; Meador *et al*., 2000; but *cf*. Bradlow and Bent, 2002: 279). This effect might allow for measurements of cognitive load and processing difficulties in various types of speech signal.

The main objective of the present study is to estimate how various features of Czech English perform in degraded speech signal. Many studies explored accent detection or rating together with intelligibility, and it has been established that evaluation of a foreign accent should not be carelessly equated with word recognition, even if they are related (Derwing and Munro, 2009; Ikeno and Hansen, 2006; Munro and Derwing, 1995). Our aim is to offer some further data on foreign accent assessment in adverse conditions independently of intelligibility and comprehensibility.

Another variable which is held constant in the present study is that of the linguistic background of the listeners. Based on the result of our previous study (Skarnitzl *et al*., 2005), we decided to use proficient Czech listeners only. This economical solution is supported by the fact that when we asked a group of native speakers of Czech and a group of native speakers of English to evaluate the degree of foreignness in Czech English (CzE) recordings, the two groups reached considerable consensus both within and mutually. No significant difference in ratings nor in the dispersion of scores was found in the data. Since our present study makes use of the recordings from the previous study, we decided to employ a homogeneous group of listeners. This decision is also in agreement with other studies where Chinese, Arabic and Japanese listeners were employed (Flege, 1988; MacKay *et al*., 2006; Munro *et al*., 2006).

Past research established that the cues indicating foreignness, i.e., departures from some specific norms, might be both segmental (consonantal or vocalic) and suprasegmental (differences in temporal organization or divergences from dynamic or melodic patterns) (e.g., Flege *et al*., 1995). There is uncertainty, however, as to how various cues of foreignness combine to contribute to the overall effect. Ingram and Pittam (1987: 128) imply that vowels reveal foreign accent better than consonants. In Magen's experiment, on the other hand, the vowel manipulations did not have an effect comparable to consonant manipulations (Magen, 1998). While Ingram and Pittam's inference could be inaccurate because they did not test the impact of consonants against vowels in any rigorous manner, Magen may have been misled by the quality of her manipulations of the speech stimuli. For instance, her attempts to introduce schwa into the speech signal apparently consisted in shortening the original full English vowel. It is not surprising that the listeners failed to find this edited form as less foreign-accented: English native schwas are not shortened peripheral vowels.

Similar controversies are mentioned by Munro and Derwing (1995: 76) in the research into the intelligibility of non-native speech: there are studies which establish the consonants as most important and other studies which claim to prove the priority of vowels or the

ACCEPTED MANUSCRIPT

suprasegmentals. Brennan and Brennan's Accentedness Index reflects 17 segmental elements but only one suprasegmental feature (Brennan and Brennan, 1981: 492). Flege (1984) showed the importance of subphonemic detail, and in his overview of L1-to-L2 influences he deals exclusively with segmental cues of foreignness (Flege, 2007). On the other hand, Magen's strongest cue of Spanish accent in English seemed to have been rhythm, namely the rhythmic abnormality caused by epenthetic vowels (Magen, 1998).

We address this issue by utilizing various types of signal degradation with some of them (e.g., low-pass filtering) affecting segmental patterns more than the suprasegmental ones. Therefore, if the suprasegmentals are largely left intact by some of the maskers, the listeners' ratings should indicate their role in foreign accent perception.

The acoustic correlates of the suprasegmental features studied in our current experiment are a) articulation rate, b) rhythm metrics, c) intensity prominence as captured by contrasts in sound pressure levels (SPL) and d) two descriptors of fundamental frequency (F0) contours.

As to articulation rate, Cooke at al. (2008: 418, 419) found that speech in noise is more intelligible if it is slower. The authors suggest that a lower speech rate may provide more processing time, thus allowing for better analysis. Similar results with speech without maskers was presented by Anderson-Hsieh and Koehler (1988). In addition, it is quite probable that lower speech rates allow for greater articulatory precision (Hazan and Markham, 2004: 3109). Given that intelligibility and perceived accentedness are to some degree correlated, it could be hypothesized that slower speakers might receive lower foreignness scores. On the other hand, MacKay et al. (2006) found that lower speech rates did not influence the impression of foreignness in a significant or consistent manner. In two of their experimental groups, faster speech seemed to be slightly better, while in the remaining two the effect was quite contrary. It is known, however, that competence is in many cultures often assessed with regard to speed: in the case of speech, fast speakers would be seen as more proficient. This is corroborated by the findings in the study by Munro and Derwing (1998), where faster speakers were perceived as more comprehensible, and again later, when the same authors manipulated the rate of speech in their stimuli and found better accentedness scores for faster items (Munro and Derwing, 2001). Our study should not only endorse some of the above-mentioned conclusions, but also investigate the robustness of articulation rate cues in acoustically adverse conditions.

Another field of our interest still lies within the time domain. The rhythm metrics are global descriptors of the temporal structure of utterances and have been shown to distinguish between rhythmically different languages (Grabe and Low, 2002; Ramus et al., 1999). They quantify durational patterns of consonantal or vocalic intervals in speech regardless of word or syllable boundaries. Despite some justified criticism and suggested improvements (Asu and Nolan, 2006; Barry et al., 2003; Gibbon and Gut, 2001; Wagner and Delwo, 2004), these parameters were found useful in prosody research and it has been shown that they can differentiate even between regional accents of one language (White et al., 2007). It is of interest to see whether they are sensitive enough to capture differences in the degree of foreign accent and if so, whether they remain sensitive under adverse conditions.

As pointed out above, the rhythm metrics are based solely on durational measures of vocalic and consonantal intervals. Apart from different temporal patterning compared to native English, we noticed that Czech speakers of English are often unable to convey the proper prominence of stressed syllables and suppress the prominence of unstressed syllables in terms of their relative loudness. We decided to take SPL measurements (in dB) of syllabic nuclei as an indirect indicator of loudness and compare the stressed nuclei with the unstressed ones and each of them with the mean SPL of the utterance. Again, the unfavourable listening conditions could map the resistance of this variable to signal degradation.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

In the fundamental frequency domain, it has been pointed out informally by native speakers of English that Czech English typically sounds 'disinterested' or even 'bored'. This impression might mean that excursions in the F0 contours of Czech English are smaller or somehow less extreme than those of native English. Moreover, we wondered whether the alleged monotony could be connected with the gradient of the general F0 declination line (cf., e.g., Paeschke and Sendlmeier, 2000; Marková *et al.*, 2009).

As to the degradation conditions, we use two types of masking noise and one type of filtering. In Experiment II, two of these degradation conditions are used at two levels – milder and harsher. In addition, our two types of noise offer a limited opportunity to test the difference between informational and energetic masking (García Lecumberri and Cooke, 2006; Cooke et al., 2008; Scott et al., 2009: 1737). Mapping the difference in the impact of stationary noise (energetic masking) and that of coffee-shop noise (energetic and informational masking) can apparently help to build more valid models of human speech perception. Bandpass filtering was opted for because of its capacity to suppress segmental, especially consonantal information and preserve the prosodic outline of utterances. We decided to work with the region of the frequency band which proved best in our earlier experiments with automatic detection of rhythm classes: it was established by trial and error that energy changes between 400 and 1400 Hz permit most successful discrimination between Czech and English read speech (Volín and Pollák, 2009).

Finally, two questions of smaller magnitude are raised in this study. One of them concerns the length of the stimuli, the other the speakers' gender. Flege (1984) showed that foreign accent can be detected in extremely short excerpts of even subsyllabic span. Some of his test items consisted of only a 30-millisecond sample of speech with subtle allophonic detail. However, his own research a decade later suggested that even sentences of about five words sometimes led to mistaken assessment of the native speakers: they were perceived by native listeners as foreigners (Flege et al., 1995). The advantage in reliability of accent detection of longer against shorter stimuli is also corroborated by Ikeno and Hansen (2006: 402). In our previous study, where we found fairly unambiguous ratings of the strength of the speakers' accent, we used stimuli of a paragraph length, i.e., from 70 to 90 words (Skarnitzl et al., 2005). These excessive samples were quite costly in terms of research time and perhaps unnecessary. Therefore, we would like to find out whether the same speakers, who received their ratings based on paragraph-long stimuli, will still get the same foreignness ratings if the stimuli consist in utterances of only 12-14 words.

The design of our study also allows for examination of whether voices of speakers of both genders produce equivalent effects in adverse conditions. Bradlow et al. (1996) and Hazan and Markham (2004) stressed that female speakers had been found more intelligible than male speakers. Gender mattered in the study of Flege et al. (1995) as well, although not in a straightforward fashion, and the authors name several other studies where the outcomes diverged from each other. The question of interest is whether the differences between genders are purely physiologically-based, i.e., the lower spectral values produced by male speaker are the chief cause of lower ratings, or whether it is the different articulatory detail: there are studies which show that males adhere less to the pronunciation standards or that they just learn to speak differently (Henton, 1990; Bladon et al., 1984 in Johnson, 2005). In order to prevent too much repetition of our test items (and the consequent scoring by recognition of the known speaker), we modified the gender of our speakers artificially (see below – Experiment II). The gender modification did not change any 'articulatory' characteristics of the speaker, thus allowing observation of the effect of plain lowering of the spectral values.

Briefly summarized, then, we would like to investigate the impact (and possible interactions) of the following variables on the assessment of foreign accent strength:

- several types of speech degradation,

ACCEPTED MANUSCRIPT

- articulation rate, rhythm characteristics, intensity contrasts and global F0 features,
- the length of the accented speech samples,
- the artificial gender change which shifts the spectral values downwards.

2. Method

2.1. Speakers

The same eight female speakers were used both in Experiment I and II. They were selected based on the results of the Skarnitzl *et al.* (2005) study: three who had been judged to sound like native or near-native speakers (henceforth Accent Group A), three speakers whose accent had been evaluated as strongly Czech (Accent Group C), and two who had been judged to display an average degree of Czech accent (Accent Group B). The speech was based on fluent reading of regular BBC news-bulletins. Three utterances were chosen per speaker, each 12 to 14 words long, with defined requirements as to the presence of specific pronunciation features typical of the Czech accent in English. As to consonantal features we required at least one velar nasal or dental fricative in each utterance. Regarding vowels at least one front open-mid or back open-mid vowel was obligatory or three phrase-internal word-initial vowels. As to the metric structure, we required at least 5 canonical schwas (mid central lax vowels), a minimum of 3 weak-form words (monosyllabic structural words), and at least one word with stress on other than the first syllable. Naturally, we did not choose any utterances with dysfluences in them.

2.2. Measurements

The key dependent variable was the mean score of a test item across all the listeners who rated it. The independent variables were categorical in the case of signal degradation and numerical in the case of articulation rate, rhythm metrics, SPL differences and F0 variability and slope. These were measured as follows.

Articulation rate was expressed in two different units following the findings of Pfitzinger (1998; also Pfitzinger and Tamashima, 2006), i.e., syllables per second and phonemes per second. Our calculations were based on canonical forms of words, which means that we counted the underlying phonemes or syllables according to the dictionary forms of words rather than actually pronounced units. This decision is based on Koreman's results which suggest that rates expressed in 'intended' phonemes are more differentiating than those expressed in realized phones (Koreman, 2006: 591).

The so-called rhythm metrics were measured according to the original proposals in Ramus et al. (1999) and in Grabe and Low (2002), and we also preserved the suggested symbols. Therefore, the percentage of vocalic portions in speech is denoted as %V, the standard deviation of consonantal interval durations as ΔC, the pairwise variability index of vocalic intervals is PVI-V and that of consonantal intervals PVI-C.

Sound pressure level (SPL) is metaphorically often called intensity in phonetic literature. It was measured in the Praat software (Boersma and Weenink, 2009). As we analyzed only female voices, the *Minimum pitch* value was raised to 140 Hz, yielding an effective window length of about 20 milliseconds. Intensity measures were acquired in the mid 30 milliseconds of all stressed and unstressed syllabic nuclei in the sentence: we calculated the arithmetic mean of five measurements 3 ms apart, centred around the mid of the nucleus. (The syllabic nucleus or peak is a phonotactic term for a vowel or syllabic sonorant.)

Fundamental frequency was measured from F0 tracks extracted by the autocorrelation method, with values taken every 10 milliseconds. The tracks were manually corrected for typical errors. The general declination trend was expressed by the gradient of the regression line through all voiced points, a method introduced by Lieberman et al. (1985).

ACCEPTED MANUSCRIPT

2.3. Material and listeners

The descriptions of the speech masking and filtering together with the testing procedures are presented in Sections 3 and 4 since the two conducted experiments required different settings.

3. Experiment I

3.1. Material

For the purposes of Experiment I, all 24 items (see Section 2.1) were mixed with two types of noise, coffeeshop noise and brown noise, at the SNR of 0 dB. (Brown noise is defined as a random noise with the spectral density of $1 / f^2$. This means that it has more energy in the low frequencies, and the spectral slope parallels the average spectral slope of natural speech). The advantage of using brown rather than white noise is in that the former masks speech more 'evenly' – it does not overemphasize high frequencies relative to speech.

For each item, the noise was faded in within 250 ms prior to the onset of speech and faded out within 150 ms from the offset of speech. As the third form of speech signal degradation, the items were bandpass-filtered between 400 and 1,200 Hz (i.e., to leave only the frequency band between 400 and 1,200 Hz).

Three analogous listening tests were designed, with each test alternately combining clean speech and one form of distortion (brown noise – BN, coffeeshop noise – CN, or filtered speech – FS). The items pronounced by the speakers who had received the best and worst scores in our previous study were repeated once, in both the clean and distorted version, yielding the total of 60 items in each test (30 clean + 30 distorted items). The items were put in a pseudo-random order, with the requirement that two utterances of the same speaker do not repeat. The items were equalized for mean intensity, which was especially important in the filtered speech utterances.

A six-second desensitization passage was used, which combined a sentence in Czech pronounced by a male voice announcing the number of each item with unobtrusive stretches of instrumental music. The total duration of the test was approximately 15 minutes.

3.2. Listeners

The three tests were administered to three groups of undergraduate students of English phonetics and phonology in a sound-treated room via high-quality loudspeakers. All the students were in their first year of studies with ages ranging between 19 and 23 years. All of them also passed an entrance examination which guaranteed their high proficiency in English. As such, they represented young educated Czech adults with very good command of English. Test I (clean speech and speech in coffeeshop noise) was administered to 24 students, Test II (clean speech and speech in brown noise) to 26 students, and Test III (clean speech and filtered speech) to 20 students. (This yielded 4200 judgements in total, on which the results of Experiment I are based.) The test was presented as an integrated part of the coursework, and the students were therefore not compensated financially for their participation.

The students were instructed to mark the degree of Czech accent in each utterance, using grades 1 to 5 on a scale which was defined similarly to Ikeno and Hansen's (2006: 402). For instance, grade 5 meant the strongest Czech accent, grade 3 represented moderate Czech English and grade 1 meant no detectable Czech accent in the given English utterance. The decision to use a five-point scale was taken for two reasons. First, it is known that the listeners' internal standards of what constitutes 'foreign accent' vary greatly (Southwood and Flege, 1999: 345). Employing the scale which is used most commonly and which Czech students are most familiar with could make our instructions and their task simpler, while avoiding some of the artifacts discussed by Southwood and Flege (1999). Second, we were

ACCEPTED MANUSCRIPT

interested principally in the midpoint of the scale and its endpoints (see above 2.1. *Speakers*). The continuous dependent variable in our analyses was built from the arithmetic means of all the scores awarded to the given speech sample by all the listeners. At this stage, more precise measurements were not necessary.

3.3. Results of Experiment I

First of all, the assumption was tested that the ratings received by our speakers in the previous study (Skarnitzl *et al*., 2005), which were based on texts of 70-90 words, are consistent with the present ratings based on items of only 12 to 14 words. To verify this assumption, the mean scores awarded to the speakers in both perception tests were entered into the Pearson's correlation formula. Since the former study did not use any signal distortions, only the items of clean speech from the current study were used. The correlation between the former scores based on long texts and the current scores was significant ($p < 0.001$) and very high: $r = 0.96$. The dispersion of the scores, however, is difficult to compare since the samples were of different magnitude (23 listeners in the former, 70 listeners in the latter study). For a rough outline, though, the mean standard deviation in the older study was 0.71 while the new study produced 0.79. The smaller dispersion of the scores in the older study may have been caused by the fact that the respondents were experienced teachers of English as a foreign language, while our current respondents were first-year university students of English.

The main concern of the present experiment was the effect of signal degradations on foreign accentedness scores. The mean scores of individual test items were entered into a mixed design ANOVA. The initial question was whether the degradation of the speech signal as such had any effect at all. There were three levels of the between-group factor (ACCENT GROUP: A, B, C) and only two levels of the within-group factor (SIGNAL DEGRADATION: no, yes). Figure 1 displays the compared means.

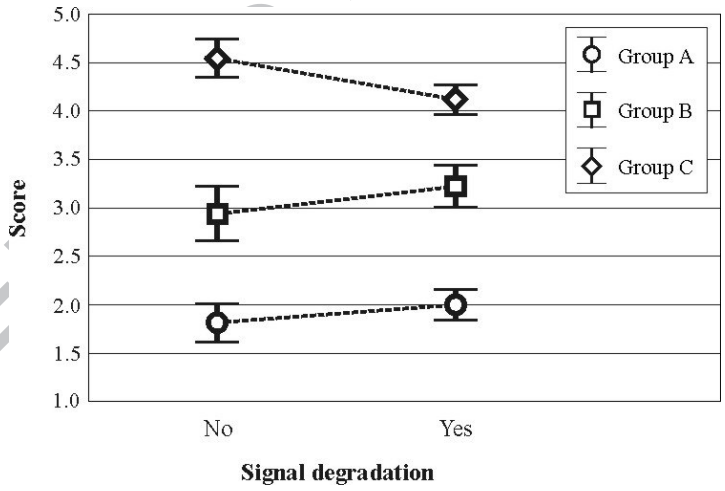

Fig. 1. Mean scores of speech items in three groups according to the strength of foreign accent. No signal degradation means clean studio recordings. The used signal degradations are described in the text. (Error bars show 95% confidence intervals.)

The analysis of variance found highly significant interaction between ACCENT GROUP and SIGNAL DEGRADATION: $F(2, 27) = 16.6$; $p < 0.001$. The post-hoc Tukey HSD tests revealed that the three accent groups stayed different from each other under both conditions, but Group C (with the strongest foreign accent) received significantly better scores in degraded speech.

ACCEPTED MANUSCRIPT

The impact of individual types of degradation was tested with a mixed design ANOVA 3×4 with three levels of the between-group factor (ACCENT GROUP: A, B, C) and four levels of the within-group factor (SIGNAL DEGRADATION: None, Brown Noise, Coffeeshop Noise, Filtered Speech). Main effects of both factors were highly significant and so was the interaction between the factors: $F(6, 81) = 29.3$; $p < 0.001$. This situation is captured in Figure 2.

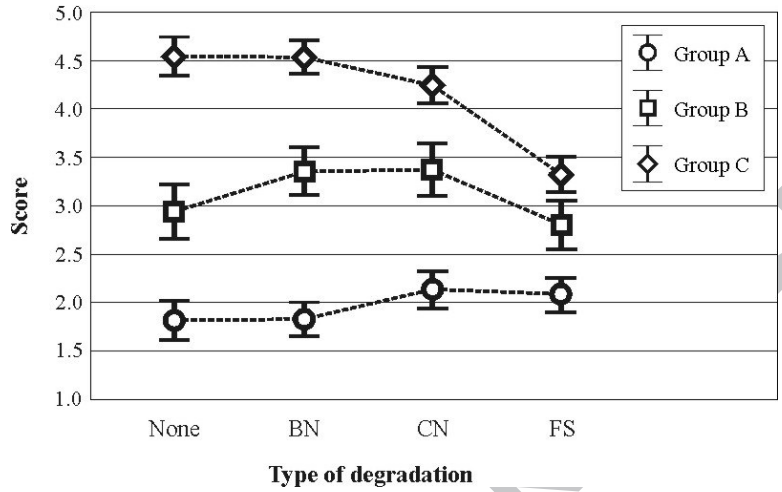

Fig. 2. Mean scores of speech items in three groups according to strength of foreign accent. (BN – brown noise, CN – coffeeshop noise, FS – filtered speech; error bars show 95% confidence intervals.)

The converging effect suggested by the previous raw analysis is better illuminated here. The post-hoc Tukey HSD tests of pairwise comparisons showed that apart from the filtered speech (FS), the three accent groups remained distinct from each other. However, the FS of Group B (moderate foreign accent) is no longer significantly different from FS of Group A (negligible foreign accent), and, more profoundly, FS of Group C (strong Czech accent) does not differ from any of the conditions of Group B. On the other hand, there are significant differences between the FS of Group B and both masked conditions of the same group, and between the FS of Group C and all the other conditions of the same group.

Furthermore, separate analysis of the difference between brown noise (BN) and coffeeshop noise (CN) produces significant worsening of Group A scores [$t(12) = -3.33$; $p < 0.01$] and significant improvement in Group C scores [$t(12) = 3.81$; $p < 0.01$]. This is ultimately a shift towards the loss of distinction among the accent groups.

Let us now cast light on phonetic properties of the individual test items and their association with the accentedness scores. First of all, we calculated correlations between the scores and articulation rates (syllabic and phonemic, see above – *Section 2.2*). Significant negative correlations were found for both types of articulation rate (AR). Apparently, faster speakers received better foreign accent evaluation. This effect is not entirely linked to the fact that speakers from Group C are slower than the other two groups. If we look at scores of individual items, in 5 cases of 8 we can see the tendency to receive better scores for faster items even when they are produced by the same speaker. Interestingly, the Pearson correlation coefficients (r) are slightly lower for the syllabic AR than for the phonemic AR (Table 1).

	Clean speech	*BN*	*CN*	*FS*
Syllabic AR	-0.59	-0.56	-0.52	-0.41
Phonemic AR	-0.61	-0.58	-0.54	-0.45

ACCEPTED MANUSCRIPT

Table 1. Pearson coefficients *r* for correlations between articulation rates (AR) and perceptual scores of foreign accentedness under different listening conditions (BN – brown noise, CN – coffeeshop noise, FS – filtered speech).

The situation is further demonstrated in Figure 3. Since the strong foreign accent is no longer perceived as strong in filtered speech, the coefficient of determination which describes the bond between articulation rate and accent score drops from about $r^2 = 0.38$ to $r^2 = 0.20$.

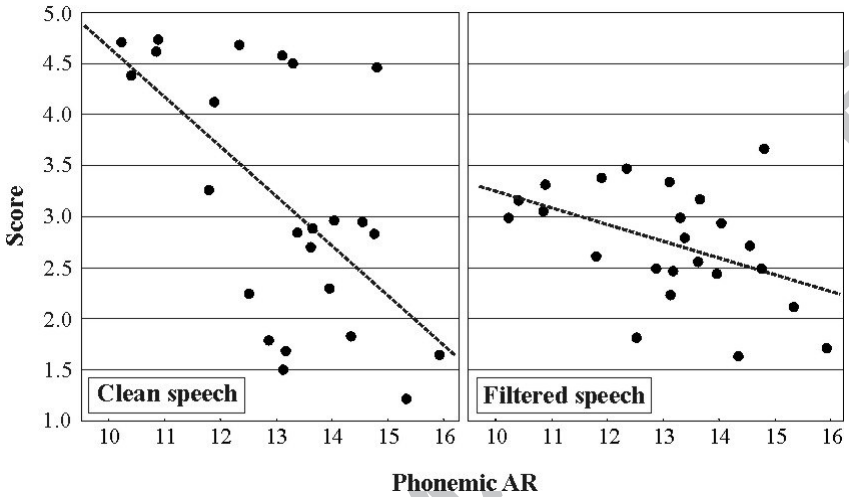

Fig. 3. Scatterplots with regression lines illustrating the relationship between articulation rate and scores of foreign accentedness.

A further set of correlation analyses concerned the so-called rhythm metrics (Grabe and Low, 2002; Ramus et al., 1999). We wanted to see whether some of these metrics correlate with foreign accent scores and if so, whether they are affected by speech signal degradation. Table 2 contains only the statistically significant results ($\alpha = 0.05$). As can be understood from their absence in the table, the consonantal rhythm metrics (ΔC and PVI-C) did not take part in any noteworthy correlations. Interestingly, the pairwise variability index of vocalic intervals (PVI-V) strengthens its correlation as the listening conditions deteriorate. The percentage of vocalic intervals within an utterance (%V) behaves conversely.

	Clean speech	*BN*	*CN*	*FS*
PVI – V	-0.46	-0.51	-0.51	-0.62
% V	-0.54	-0.55	-0.54	-0.45

Table 2. Pearson coefficients *r* for correlations between two rhythm metrics and perceptual scores of foreign accentedness (BN – brown noise, CN – coffeeshop noise, FS – filtered speech).

Figure 4 demonstrates the situation for the clean signal (no speech degradation). In parallel with our previous findings (some of which were published in Volín and Pollák, 2009), it is obvious that the greater the portion of vowels in the duration of the utterance, the more English and less Czech the speech sounds. Similarly, the higher the vowel duration variation, the more English and less Czech the speech should be.

ACCEPTED MANUSCRIPT

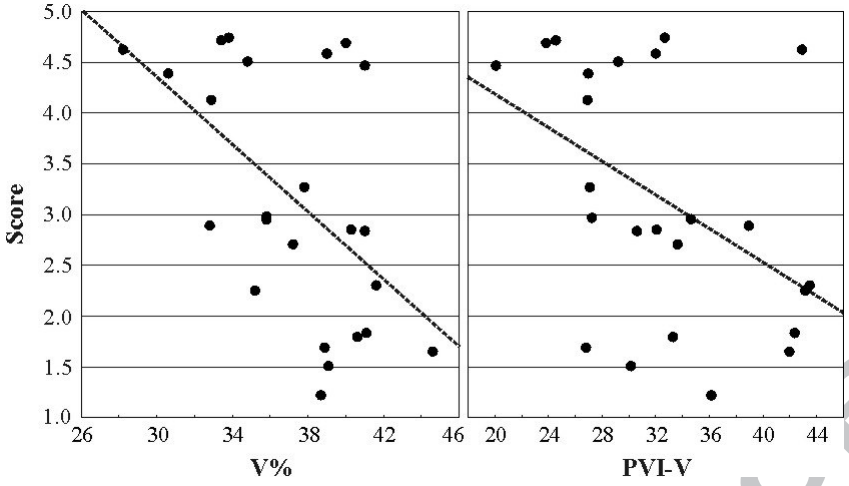

Fig. 4. Scatterplots with regression lines illustrating the relationship between the global rhythm metrics (for %V, PVI-V see text) and scores of foreign accentedness.

The third set of correlations concerned SPL in stressed and unstressed syllables. We measured a) the difference between means of stressed and unstressed nuclei, b) the difference between stressed nuclei and the overall mean of the given utterance, and c) the difference between unstressed nuclei and the overall mean of the given utterance. The correlation coefficients are displayed in Table 3. Clearly, the difference between stressed and unstressed syllables has the greatest explanatory value. The differences from the mean are probably blurred by the inclusion of voiceless consonants into the mean, which makes the measurements dependent on the segmental composition of the utterance.

	Clean speech	*BN*	*CN*	*FS*
stress – unstress	-0.58	-0.56	-0.53	-0.51
mean – stress	-0.52	-0.45	n.s.	n.s.
mean – unstress	n.s.	n.s.	-0.45	-0.49

Table 3. Pearson coefficients *r* for correlations between three SPL parameters and perceptual scores of foreign accentedness (BN – brown noise, CN – coffeeshop noise, FS – filtered speech; significance at $\alpha = 0.05$).

The situation is further illustrated in Figure 5, which also reveals that some of the speakers with a strong Czech accent produced unstressed syllables with a higher mean SPL than the stressed ones. Contrary to that, speakers with a negligible Czech accent had stressed syllables up to 3 dB stronger than the unstressed ones.

An interesting comparison is offered especially by the individual items of the same linguistic content (the identical utterances pronounced by different speakers). For instance, item 7 received a surprisingly poor mean score despite the fact that it came from Accent Group A. Actually, it was the only item from Group A to fall into the Group B range of scores, even though it did not contain any obvious deviations from native-like pronunciation on the segmental level. As it was based on the same sentence as item 4, direct comparison was possible. The speaker who achieved better scores produced the stressed syllables stronger than the unstressed ones by 3.2 dB. For the worse scores the difference between stressed and unstressed syllables was less than 0.4 dB. Similarly, item 3 was produced by yet another good speaker with the difference between stressed and unstressed syllables of 3.3 dB, while the

ACCEPTED MANUSCRIPT

same text uttered by a Group C speaker produced a difference of only 0.27 dB and was consistently awarded poor scores.

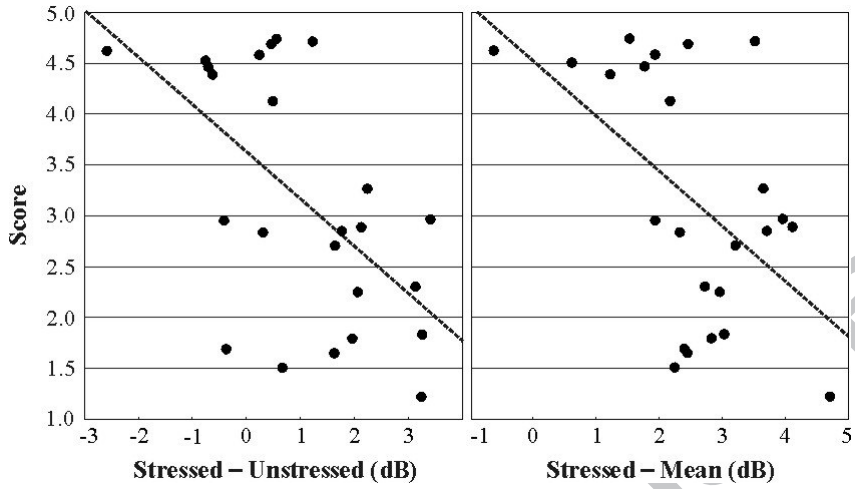

Fig. 5. Scatterplots with regression lines illustrating the relationship between the SPL measures and scores of foreign accentedness in clean speech.

Finally, the F0 variation and declination produced the correlation coefficients displayed in Table 4. The variation is captured in the standard deviation of voiced points (taken every 10 ms) from the mean and in the percentile range: the distance between the 10th and the 90th percentile. The declination slope was expressed by the gradient of the regression line in semitones per second through all the voiced points in the F0 track.

	Clean speech	*BN*	*CN*	*FS*
Std. dev. (F0)	-0.61	-0.63	-0.59	-0.65
Perc. range (F0)	-0.61	-0.62	-0.58	-0.63
Slope (F0)	-0.61	-0.60	-0.50	-0.51

Table 4. Pearson coefficients *r* for correlations between three F0 parameters and perceptual scores of foreign accentedness (BN – brown noise, CN – coffeeshop noise, FS – filtered speech; significance at $\alpha = 0.05$).

The correlation coefficients suggest that the greater the variation in the F0 contour, the better accentedness scores can be expected. The general declination trend behaves in the same manner. The difference between the two phenomena is that variation in F0 contours seems to be robust to signal degradation, while the slope of declination loses somewhat its importance in adverse conditions.

4. Experiment II

4.1. Material

Given the results of Experiment I, we wanted to collect some further data on its more severe listening conditions (CN and FS). Clean speech proved to be rated with satisfactory consistency by listeners of various backgrounds (more experienced × less experienced) and in different speech samples (long × short). Stationary brown noise at the SNR of 0 dB did not produce any significant differences from the clean speech ratings. Therefore, we decided to

ACCEPTED MANUSCRIPT

"stretch" the more adverse conditions: the coffeeshop noise (also because it better corresponds to real-life listening situations) and filtering (also because it suppresses segmental information in the speech signal while leaving suprasegmental features largely intact).

Of the original 24 items, only 14 were selected for Experiment II. First of all, we wanted to retain the items with greatest similarity in the linguistic content. Second, we excluded the items which produced less consistent results in Experiment I. The accent groups were represented as follows:

Group A – speakers with a negligible Czech accent: 4 items,
Group B – speakers with a moderate Czech accent: 4 items,
Group C – speakers with a strong Czech accent: 6 items.

Each of these 14 items, pronounced by female voices, was converted to a male voice using PSOLA, specifically the *Change gender* function in Praat (Boersma & Weenink, 2009). F0 median was reduced by 9 semitones, and the formant ratio was specified as 0.85 for the conversion. The new items were checked by both authors for possible errors resulting from the conversion. Eventually, there were 28 source items: 14 produced by female voices + 14 of converted male voices. In terms of coffeeshop noise, the signal-to-noise ratio of 0 dB (C-I) was complemented with a more severe masking of -6 dB (C-II). As to the bandpass-filtering, we decided to add a more moderate distortion of the speech signal with a wider band of frequencies: 400-1,400 Hz (F-I), and preserved the more severe one with the specification of 400-1,200 Hz (F-II) from the previous experiment. In the following presentation of results, these conditions will be ordered according to the predicted severity: C-I, F-I, F-II and C-II.

Two versions of the listening test were compiled, with one version including only either the original female or the resynthesized male version of a given item. In fact, the two versions were "mirrors", differing only in the female-male characteristic for each item. The items were again equalized for mean SPL and put in a pseudo-random order. Similarly to the previous experiment, a six-second desensitization passage with a Czech male voice and unobtrusive stretches of instrumental music was used to counter the effect of ordering.

4.2. Listeners

Both versions of the test were administered to 26 students who had taken the English phonetics and phonology course in previous academic years and volunteered to participate in the experiment. None of them had thus participated in Experiment I and none of them reported any speech or hearing impairments. The tests were administered individually via the KOSS UR/15 headphones. The approximately 15-minute test was divided into two halves, with a 5-minute break between them during which the subjects chatted with the administrator of the test to avoid perception fatigue. The two sessions, each for a different version of the test, were separated by at least one week.

4.3. Results of Experiment II

Twenty-six listeners assessed 56 test items in each of the two sessions. Therefore, the following results should be based on 2,912 judgements (26 × 56 × 2). However, 34 of the judgements were not delivered. In the follow-up interview, some of the listeners reported complete lack of confidence at a few items; those had subsequently not been assessed. All but two of the missing scores belonged to the most severe condition C-II (coffeeshop noise with SNR = -6dB). The results of Experiment II are therefore based on 2,878 judgements. Moreover, two of the listeners did not use the whole scoring range when evaluating the test items. Their judgements were rescored to cover the scoring range from 1 to 5.

The first question to answer was that of the comparability of both experiments. Masking with coffeeshop noise at the SNR of 0 dB (C-I) and filtering with the bandpass of 400-1,200 Hz (F-

ACCEPTED MANUSCRIPT

II) were played to the listeners in the current experiment individually via headphones, while the listeners in Experiment I heard them in a sound-treated room through loudspeakers in groups of up to 12 students. Repeated measures ANOVA with the Tukey HSD post-hoc tests showed that four out of six comparisons were not significantly different, but there was a significant difference between the scores of Group B for C-I masking and in Group C for filtering F-II. The direction of these differences suggests that headphones make the listening conditions somewhat less adverse. They slightly reduce the converging effect found in Experiment I (see Figure 2) in that the listeners with headphones better discriminate among the three accent groups relative to the clean speech scores.

The mean scores for all four conditions in Experiment II were first tested by a simple ANOVA for repeated measures. This revealed a significant effect of SIGNAL DEGRADATION: $F(3, 81) = 5.34$; $p < 0.01$. Post-hoc Tukey HSD test singled out the difference between C-II masking and F-II filtering as the only significant contrast. However, these raw results are just exploratory and potentially even misleading as they collapse all the three accent groups together, which a) unnecessarily increases the standard error of the scores and b) hides the possible non-parallel effects in individual accent groups.

Therefore, a mixed design ANOVA 3×4 was performed with ACCENT GROUP as the between-group variable (levels A, B, and C) and SIGNAL DEGRADATION as the within-group variable. The situation is captured in Figure 6.

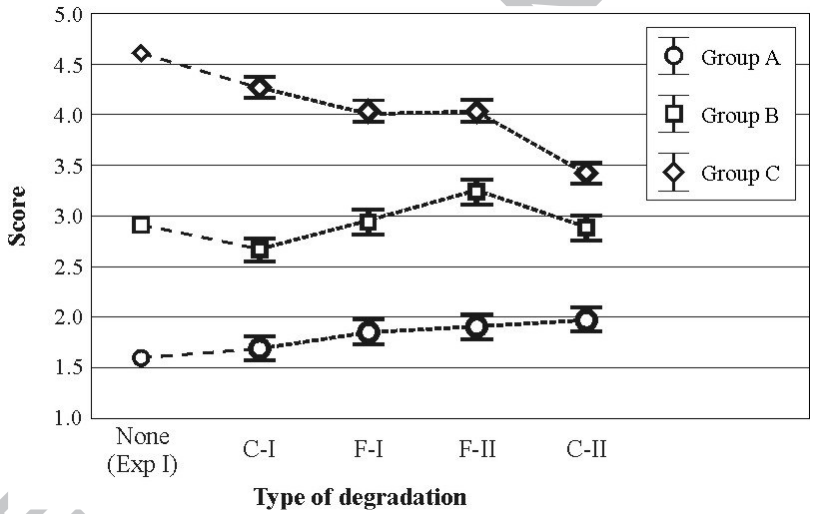

Fig. 6. Mean scores with 95% confidence intervals for the four speech degradation conditions in three accent groups according to foreignness in their English.

Both main effects and interactions were significant [$F(2, 693) = 761.7$, $p < 0.001$; $F(3, 2079) = 19.5$, $p < 0.001$; $F(6, 2079) = 34.0$, $p < 0.001$, respectively].

Figure 6 also displays the means for clean speech from Experiment I for comparison, but those scores were not used in the ANOVA calculations. The figure confirms the trend suggested in Experiment I. The deteriorating conditions decrease the differences between accent groups. Post-hoc Tukey HSD tests confirmed significant differences among the conditions in Accent Group C except between F-I and F-II. On the other hand, it suggested confusion between the C-II condition in Group C and F-II in Group B. The differences among conditions in Group B also reached significance apart from the one between F-I and C-II. The only significant difference in Group A is the one between C-I and C-II, i.e., between the softest and the most severe listening conditions.

ACCEPTED MANUSCRIPT

The correlation analysis of foreign accent scores and selected suprasegmental indicators also confirms the findings from Experiment I, but adds some new details (Table 5). As to articulation rate, the phonemic AR remains superior to the syllabic AR in that it yields higher correlation coefficients with foreign accent scores.

	C-I	*F-I*	*F-II*	*C-II*
Syllabic AR	-0.61	-0.55	-0.58	-0.46
Phonemic AR	-0.67	-0.61	-0.62	-0.49
PVI-V	-0.41	-0.54	-0.53	-0.52
% V	-0.41	-0.43	-0.41	n.s.
stress – unstress (SPL)	-0.61	-0.71	-0.63	-0.59
Std. dev. (F0)	-0.76	-0.90	-0.91	-0.92
Slope (F0)	0.68	0.82	0.78	0.81

Table 5. Pearson coefficients *r* for correlations between articulation rates, rhythm metrics, SPL and F0 parameters and perceptual scores of foreign accentedness under different listening conditions (significance at $\alpha = 0.05$).

The rhythm metrics are again represented by the vocalic measures only. The consonantal parameters did not yield significant coefficients. Pairwise variability index (PVI-V) and F0 standard deviation replicate the behaviour from Experiment I, this time also accompanied by the slope of the declination line. Whereas most of our suprasegmental measures produce weaker correlations under less favourable conditions, these three measures do the opposite: the weakest correlations are found under the most favourable listening condition, C-I.
Finally, we wanted to know if the artificial change of gender characteristics in our samples affected the listeners' judgements in any consistent manner. Figure 7 indicates that while on average the original female and the modified male items received about the same scores, there were differences in individual accent groups. A mixed design ANOVA 3×2 with ACCENT GROUP as the between-group variable (levels A, B, and C) and GENDER as the within-group variable (Female and Modified Gender) was performed with the significant interaction between the factors: $F(2, 44) = 7.72$; $p = 0.0013$. The post-hoc t-test for dependent samples revealed a significant difference for Groups A and C. Modified gender yielded worse results in the group of the best speakers [$t(15) = 2.75$; $p = 0.015$] and better results in the group of the speakers with a strong accent [$t(23) = –2.61$; $p = 0.016$. This result parallels that of other types of the speech signal degradation, which also caused perceptual worsening in Group A and improvement in Group C.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

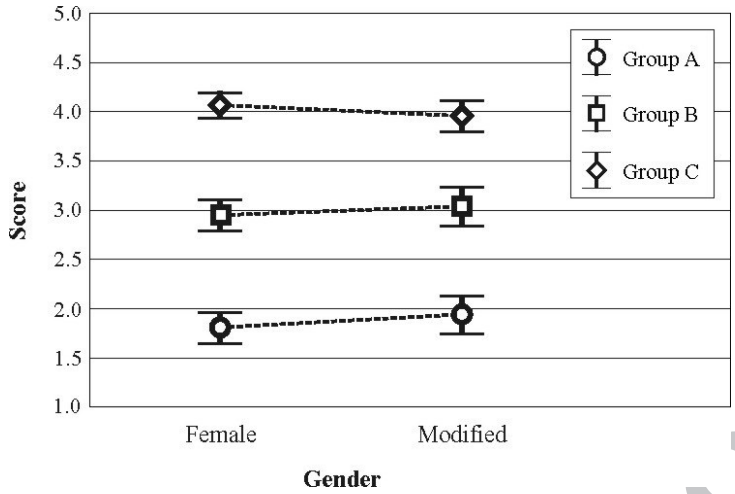

Fig. 7. Mean scores of the perceived foreign accent strength according to gender and the proficiency of the speaker.

5. Discussion and Conclusion

High correlation between the foreign accent scores from our 2005 study and the current evaluations ($r = 0.96$) confirms that the Czech accent of English is perceived by various listeners under differing conditions very consistently. This is in line with Hazan and Markham's conclusion (2004) that the differences between listeners may be less important than the differences between speech samples. Contrary to that, Flege et al. (1995: 3132) reported substantial individual differences among listeners from seemingly homogeneous groups and cautioned against a careless assumption that human perception of speech is uniform. Bent and Bradlow's results (2003: 1608) speak similarly. Given the outcomes of our experiments, we would like to corroborate both stances. On the one hand, there are almost identical *mean* ratings based on different speech samples and different respondent groups. On the other hand, three of our less experienced listeners were insensitive to the strength of the foreign accent (they were near "accent-deaf"). These listeners entered our ANOVA calculations in the form of increased standard deviations from the means.

Our chief concern in the present study was the perception of foreign accentedness under various adverse conditions, and robustness of various suprasegmental cues in degraded speech signal. The most general, though not unexpected, finding from both of our experiments was a gradual equalization of the accent differences with deteriorating conditions: good speakers received increasingly worse scores, while poor speakers were assessed as progressively better. The perceptual uncertainty caused by adverse conditions does not translate in either an ameliorating or aggravating effect across the board. Rather, it pulls towards the magnet of the mid value. The question for future research remains whether this is a solely perceptual effect (which is most likely) or whether it might have a cultural (attitudinal) foundation.

A more specific finding from our Experiment I concerned the difference in masker types. Brown noise (BN) at the SNR of 0 dB did not actually produce any significant differences from clean speech, while coffeeshop noise (CN) at the same level compromised foreign accent perception more. This is in agreement with the findings of García Lecumberri and Cooke (2006), whose multi-talker babble was more challenging for the listeners than stationary noise. Apparently, energetic masking is easier to deal with than informational masking even in the case of foreign accent evaluation and not only in the case of linguistic unit identification as in García Lecumberri and Cooke's study. It is quite probable that the

ACCEPTED MANUSCRIPT

listener's brain employs different circuits according to the type of masking, and some of the circuits are more exacting than others (e.g., Scott et al., 2009: 1742).

Both Experiment I and II confirmed moderate to high correlations ($0.4 < r < 0.9$) of foreign accent scores with variables reflecting suprasegmentals of speech. Faster test items were perceived more native-like, suggesting that proficiency and speed often go hand in hand. This is in agreement with the findings of Munro & Derwing (2001), but in contrast with the intelligibility research, where it is the lower articulation rates that allow for better recovery of lexical items.

The correlation between foreign accent ratings and the pairwise variability index of vocalic intervals (PVI-V) also makes sense and so does the SPL difference between stressed and unstressed syllabic nuclei. The more variation in durations of vocalic intervals and in stressed/unstressed oppositions there is, the more native-like the speech sounds. This is a relatively direct reflection of the Czech prosodic patterns which, in comparison with English, display smaller differences in energy between individual syllables (see e.g., Volín and Pollák, 2009).

The behaviour of F0 tracks also confirms informal observations of Czech and English intonational patterns. Greater variation, i.e., greater contrastivity of English melodies in terms of local range changes is quite apparent, although, to our best knowledge, not described in phonetic literature. What was unexpected was the role of the downtrend in F0 contours reflected by the gradient of the all-point regression line. Our experiments show that less accented Czech English produces greater downtrends. This, again, is information which we have not encountered in literature on the subject matter.

We identified three measures which improve their prediction power with the increase in speech signal degradation. Whereas the other suprasegmental measures reach weaker correlations under less favourable conditions, PVI-V, F0 standard deviation and partially also gradient of F0 slope do the opposite. Their strongest correlations are found under the most adverse listening conditions. This effect should be explored more thoroughly in the near future and hopefully find its applications in foreign accent detection in ASR systems.

A much weaker parallel can be observed in the case of articulation rates and filtered speech in Experiment II: the more moderate filtering (F-I) led to slightly lower correlations than the narrower bandpass filter (F-II). However, in this case the differences are too small to allow for any generalization. Perhaps an experiment dedicated specifically to this problem could reveal more.

Our final remark concerns the interplay between segmental and suprasegmental phenomena in speech. The filtering in Experiment I constituted the most severe listening condition. The FS of Group B no longer differed in its accentedness scores from either A or C: the three groups virtually merged (although A remained different from C as an opposite tail of the score distribution). Obviously, if it were just the raw rhythmic layout that underlies the foreignness, the drop in differentiation among accent groups would not be so steep, since this raw rhythm was preserved in filtered speech. The rhythm of speech is apparently perceived relative to some fine segmental detail and not in its raw form abstracted from the more minute segmental features.

In conclusion, then, the design of our experiments allows us to hypothesize that in real life situations, which are notoriously acoustically adverse, some of the suprasegmental features preserve the perception of foreignness in speech. We have detected several candidates that are worth pursuing in further research (articulation rate, rhythm parameters, intensity and F0 contrasts and a correlate of intonation declination) and will complement them with attempts to quantify segmental departures peculiar to Czech English.

Acknowledgements

ACCEPTED MANUSCRIPT

We would like to thank our anonymous reviewers for many helpful comments on earlier versions of this paper. This work was supported by the European Union grant MRTN-CT-2006-035561 – *Sound to Sense*, and by the Czech Ministry of Education grant VZ MSM0021620825.

References

Anderson-Hsieh, J. Koehler, K., 1988. The effects of foreign accent and speaking rate on native speaker comprehension. Language Learning 38/4, 561-613.
Asu, E.L., Nolan, F., 2006. Estonian and English rhythm: a two-dimensional quantification based on syllables and feet, in: Proceedings of Speech Prosody, Dresden, pp. 249-253.
Barry, W.J., Andreeva, B., Russo, M., Dimitrova, S., Kostadinova, T., 2003. Do rhythm measures tell us anything about language type?, in: Proceedings of 15th ICPhS, Conference Organizers, Barcelona, pp. 2693-2966.
Bent, T., Bradlow, A.R., 2003. The interlanguage speech intelligibility benefit. J. Acoust. Soc. Am. 114/1, 1600-1610.
Bladon, R.A., Henton C.G., Pickering, J.B., 1984. Towards an auditory theory of speaker normalization. Language Communication 4, 59-69.
Boersma, P., Weenink, D., 2009. Praat: doing phonetics by computer (Version 5.1). Retrieved Feb 10, 2009, from http://www.praat.org/.
Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Am. 112/1, 272-284.
Brennan, E.M., Brennan, J.S., 1981. Measurements of accent and attitude toward Mexican-American speech. Journal of Psycholinguistic Research 10/5, 487-501.
Cooke, M., García Lecumberri, M.L., Barker, J., 2008. The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. J. Acoust. Soc. Am. 123/1, 414-427.
Derwin, T.M., Munro, M.J, 2009. Putting accent in its place: Rethinking obstacles to communication. Language Teaching 42/4, 476-490.
Eskenazi, M., 2009. An overview of spoken language technology for education. Speech Communication 51, 832-844.
Flege, J.E., 1984. The detection of French accent by American listeners. J. Acoust. Soc. Am. 76, 692-707.
Flege, J.E., 1988. Factors affecting degree of perceived foreign accent in English sentences. J. Acoust. Soc. Am. 84, 70-79.
Flege, J.E., 2007. Language contact in bilingualism: Phonetic system interactions, in: Cole, J., Hualde, J.I. (Eds.), Laboratory Phonology 9, Mouton de Gruyter, Berlin, pp. 353-382.
Flege, J.E., Munro, M.J., MacKay, I.R.A., 1995. Factors affecting strength of perceived foreign accent in a second language. J. Acoust. Soc. Am. 97/5, 3125-3134.
García Lecumberri, M.L., Cooke, M., 2006. Effect of masker type on native and non-native consonant perception in noise. J. Acoust. Soc. Am. 119/4, 2445-2454.
Ghesquiere, P., Van Compernolle, D., 2002. Flemish accent identification based on formant and duration features, in: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Vol. I, pp. 749-752.
Gibbon, D., Gut, U., 2001. Measuring speech rhythm, in: Eurospeech Proceedings, ISCA, Aalborg, pp. 91-94.
Grabe, E., Low, E.L., 2002. Durational variability in speech and the rhythm class, in: Gussenhoven, C., Warner, N. (Eds.), Papers in Laboratory Phonology 7, Mouton de Gruyter, Berlin, pp. 515-546.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Hahn, L.D., 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. TESOL Quarterly 38/2, 201-223.
Hazan, V., Markham, D., 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. J. Acoust. Soc. Am. 116/5, 3108-3118.
Henton, C., 1990. One vowel's life (and death?) across languages: the moribundity and prestige of /ʌ/. Journal of Phonetics 18/2, 203-228.
Ikeno, A., Hansen, J.H.L., 2006. Perceptual recognition cues in native English accent variation: Listener accent, perceived accent and comprehension, in: Proceedings of ICASSP 2006, Conference Organizers, Toulouse, pp. 401-404.
Johnson, K., 2005. Speaker normalization in speech perception, in: Pisoni, D.B., Remez, R.E. (Eds.), The Handbook of Speech Perception, Blackwell, Oxford, pp. 363-389.
Koreman, J., 2006. Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. J. Acoust. Soc. Am. 119/1, 582-596.
Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., Miller, M., 1985. Measures of the sentence intonation of read and spontaneous speech in American English. J. Acoust. Soc. Am. 77, 649-657.
Magen, H.S., 1998. The perception of foreign-accented speech. Journal of Phonetics 26, 381-400.
MacKay, I.R.A., Flege, J.E., Imai, S., 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. Applied Psycholinguistics 27, 157-183.
Marková, P., Filippová, E., Volín, J., 2009. Suprasegmental markers of irony in acted and natural discourse, in: Proceedings of XIXth Czech-German Workshop Speech Processing – Prague, IPE ASCR, pp. 56-60.
Mayo, L.H., Florentine, M., Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. J. Speech Lang. Hear. Research 40, 686-693.
Meador, D., Flege, J.E., MacKay, I.R.A., 2000. Factors affecting the recognition of words in a second language. Bilingualism: Language and Cognition 3, 55-67.
Munro, M.J., Derwing, T.M., 1995. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. Language Learning 45/1, 73-97.
Munro, M.J., Derwing, T.M., 1998. The effects of speech rate on the comprehensibility of native and foreign-accented speech. Language Learning 48/2, 159-182.
Munro, M.J., Derwing, T.M., 2001. Modelling perceptions of the accentedness and comprehensibility of L2 speech. Studies in Second Language Acquisition 23, 451-468.
Munro, M.J., Derwing, T.M., Morton, S.L., 2006. The mutual intelligibility of L2 speech. Studies in Second Language Acquisition 28, 111-131.
Paeschke, A., Sendlmeier, W.F., 2000. Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements, in: Proceedings of the ISCA ITRW on Speech and Emotion, Newcastle, 5-7 September 2000, Belfast, Textflow, pp. 75-80.
Pfitzinger, H.R., 1998. Local speech rate as a combination of syllable and phone rate, in: Proceedings of 5th ICSLP, ISCA, Sydney, pp. 1087-1090.
Pfitzinger, H.R., Tamashima, M., 2006. Comparing perceptual local speech rate of German and Japanese speech, in: Proceedings of 3rd Speech Prosody, vol. 1, TUD, Dresden, pp. 105-108.
Pickering, L., 2001. The role of tone choice in improving ITA communication in the classroom. TESOL Quarterly 35/2, 233-253.
Ramus, F., Nespor, M., Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73, 265-292.
Rubin, D.L., 1992. Non-language factors affecting undergraduate's judgments of non-native English speaking teaching assistants. Research in Higher Education 33/4, 511-531.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Scott, S.K., Rosen, S., Beaman, C.Ph., Davis, J.P., Wise, R.J.S., 2009. The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. J. Acoust. Soc. Am. 125/3, 1737-1743.
Skarnitzl, R., Volín, J., Drenková, L., 2005. Tangibility of foreign accents in speech: the case of Czech English, in: Grmelová, A. Dušková, L., Farrell, M. (Eds.), 2nd Prague Conference on Linguistics and Literary Studies: Proceedings, UK PedF, Praha, pp.11-20.
Southwood, M.H., Flege, J.E., 1999. Scaling foreign accent: direct magnitude estimation versus interval scaling. Clinical Linguistics & Phonetics 13/5, 335-349.
Strik, H., 2003. Speech is like a box of chocolates, in: Proceedings of 15th ICPhS, Conference Organizers, Barcelona, pp. 227-230.
Volín, J., Pollák, P., 2009. The dynamic dimension of the global speech-rhythm attributes, in: Proceedings of 10th Interspeech, ISCA, Brighton, pp. 1543-1546.
Wagner, P., Dellwo, V., 2004. Introducing YARD and re-introducing isochrony to rhythm research, in: Proceedings 2nd Speech Prosody, SProSIG, Nara, pp. 227-230.
White, L., Mattys, S., Series, L., Gage, S., 2007. Rhythmic metrics predict rhythmic discrimination, in: Proceedings of 16th ICPhS, Conference Organizers, Saarbrücken, pp. 1009-1012.
Wu, T., Dochateau, J., Martens, J-P., Van Compernolle, D., 2010. Feature subset selection for improved native accent identification. Speech Communication 52, 83-98.

ACCEPTED MANUSCRIPT