

# SPEECH MELODY PROPERTIES IN ENGLISH, CZECH AND CZECH ENGLISH: REFERENCE AND INTERFERENCE

*JAN VOLÍN*

Metropolitan University, Prague  
jan.volin@ff.cuni.cz

*KRISTÝNA POESOVÁ*

Charles University in Prague  
kristyna.poesova@pedf.cuni.cz

*LENKA WEINGARTOVÁ*

Institute of Phonetics, Prague  
lweingartova@gmail.com

## Abstract

Two major objectives were set for the present study: to provide reference data for the description of Czech and English F0 contours, and to investigate the limits of the ‘interference hypothesis’ on Czech English data. Altogether, the production of 40 speakers in 2392 breath-group F0 contours was analyzed. The speech of 32 professional speakers of English and Czech provides reference values for various acoustic correlates of pitch level, pitch span and downtrend gradient. These values were subsequently used as a benchmark for a confirmation of the interference hypothesis through comparison with a further sample of 8 non-professional speakers of English and Czech-accented English. The native English speakers of both genders produced significantly higher pitch level indicators, wider pitch span and a steeper downtrend gradient than the reference native speakers of Czech. Although the pitch level of the Czech-accented material lies in between the two reference groups, the pitch span of this group is the narrowest, which indicates that factors of foreign-accentedness other than simply interference are in effect.

## 1. Introduction

Speech melody fulfils a number of important functions in the sound system of languages. For instance, it facilitates the expression of interpersonal stances or feelings towards the dialogue partner or the subject of conversation. It can also signal the grammatical structure of the spoken material as well as indicate utterance type and resolve syntactic ambiguities. Furthermore, pitch variation, which is central to intonation, is commonly employed to highlight important lexical constituents in a given utterance. Last but not least, intonation helps to organise and regulate conversational behaviour and mark personal and social identity (Wells, 2006). While English relies predominantly on intonational cues when expressing contrastivity and signalling major information, other languages (including Czech) may exploit different linguistic vehicles

such as grammatical inflection or word order for the same purposes (Rogerson-Revell, 2011) and they often do so. This discrepancy may partly clarify the findings of a small-scale research the topic of which was evaluation of foreign learners' beliefs about English intonation. The results indicated strong unawareness of multiple uses of English intonation in spoken discourse even among advanced users of English. Specifically, it was reported that students tend to grasp the meaning exclusively from words. They may not believe that intonation affects meaning and therefore they are likely to view its role as rather decorative (Gilbert, 2014; Reed, 2015).

The above mentioned misconception about English intonation, which, we presume, would be similarly widespread in the Czech educational context, had it been assessed, cannot be successfully dispelled without effective pedagogical interventions based on systematic phonetic research. Students need to get a clearer picture of the fact that intonation not only reinforces or mitigates spoken words but it frequently trumps them in the sense that a particular tone and/or pitch choice can override the meaning expressed on the lexical level (Wichmann, 2005). Besides this, the significance of intonation may be mediated by showing authentic examples in which inappropriate use of pitch movement may result in communication disruption or even breakdown. Naturally, the whole area is extraordinarily complex, as intonation displays high variability, its individual functions overlap and speech melodies cannot be completely divorced from the interplay of other prosodic dimensions, e.g., patterns of timing, loudness, voice quality, or speech rate (Nolan, 2006). These complexities may lead to the lack of teacher confidence or at times even reluctance to address intonation-related issues in their classrooms which might mirror in misleading standpoints of the students themselves.

A powerful argument in favour of incorporating intonation into language classes is the fact that both pronunciation experts and phoneticians have repeatedly confirmed that suprasegmental phenomena largely contribute to achieving comfortable intelligibility, which has been the goal of pronunciation teaching and learning for the past three decades (Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing, Munro, & Wiebe, 1998; Munro & Derwing, 2005). For example, in the study dealing with Japanese-accented French, native-like prosody improved the evaluation of naturalness of utterances with non-native like segments. The perception experiments carried out with French native listeners employed synthesized stimuli, changing the segment quality and keeping the prosody unmodified in the first test and manipulating F0 and duration and keeping the segments intact in the second test (Kamiyama, 2004). Similar methodology was applied in the intelligibility assessment of Norwegian produced by speakers coming from seven different mother-tongue backgrounds. Their utterances with superimposed native-like intonation contours obtained higher intelligibility scores in almost all L1 languages, although only for English and German speakers the effect of manipulated intonation turned out to be statistically significant (Holm, 2007). Hahn (2004) used a different experimental design and found that the main variable, proper nucleus placement, facilitated comprehension and speeded up processing of non-native English to a larger extent than when the target feature was misplaced or entirely missing.

Undoubtedly, intonation along with other prosodic aspects plays a key role in the improvement of intelligibility. However, questions such as which intonation errors compromise meaning most severely and whether these errors are influenced more by universal patterns in second language acquisition or by transfer from L1s, remain

unresolved. Mennen (2007, p. 55) offers a non-hierarchical list of recurrent non-native intonation problems collected from a survey of approximately fifteen production-oriented L2 intonation studies with the aim of establishing differences and similarities across languages in the area of speech melody. One of them, a narrower pitch range, comprises the theme of the current paper. Our first objective accords with Mennen's cross-linguistic investigation as we seek to verify the interference hypothesis through comparison of native Czech and native English data with the data extracted from Czech English speech production. Thus, we hope to contribute to the existing knowledge about the reduction of pitch range in L2 acquisition with a so far unexplored language sample. The second research strand addresses the popular beliefs about the melody of Czech-accented English, which typically sounds flat and monotonous to both native and proficient non-native ears, as if signalling boredom, disinterest or lack of involvement. We aim to reveal the true nature of these perceptual impressions using rigorous descriptors of the underlying F0 tracks and see if they support this widely held view.

Pitch ranges are believed to vary across languages and there is growing evidence that even in the absence of physiological differences linguistic communities tend to be characterized by particular pitch profiles (Dolson, 1994). Again, our knowledge in this respect is fragmentary, yet interesting. Andreeva and her colleagues (2014) investigated pitch range and variation in two Germanic and two Slavic languages. The groups turned out to be convincingly distinct with English and German speakers using a considerably lower level, narrower span, and generally less variable pitch than their Bulgarian and Polish counterparts. F0 measures in this study did not indicate any significant difference within the Germanic family which corroborates neither stereotypical observations, nor Mennen's findings (Mennen, Schaeffler, & Docherty, 2007). She and her colleagues identified a wider pitch span but not pitch level for standard British female speakers compared to women speaking standard Northern German. The possibility that the perception of cross-language differences in span may be affected by other factors was not ruled out (*ibid.*). The assumption about the existence of language specific F0 profiles was also confirmed for Mandarin versus English productions. Nevertheless, the authors warn against the influence of limited speech material on the obtained differences (Keating & Kuo, 2010).

Our research goals include the establishment of F0 profiles for Czech and English read speech. We believe that data provided by 32 professional speakers may serve as *reference values* beyond our current study. However, our primary concern here lies in L2 acquisition processes and the phonetic reality of Czech-accented English. The main research question concerning this subject matter is as follows: Does the interference theory hold if we compare native Czech, English and non-native Czech English F0 tracks? In other words, do the Czech English parameters lie between native Czech and English values? The null hypothesis would propose identical values for all three modes (Czech, English and Czech English). Alternative hypotheses should not rule out any potential differences.

## 2. Method

Our core material consisted of news bulletins from the Czech National Radio read by 8 male and 8 female Czech professional news readers, and BBC news bulletins read, likewise, by 16 English professionals (half males and half females) to obtain the benchmark F0 measures for read speech in Czech and English. The acquired material amounted to about 2 hours of speech (4-4.5 min per speaker) and is referred to as the *reference condition*. The sample was further expanded by recording and analysing a set of 112 English sentences read by 4 Czech and 4 English non-professional speakers. The Czech respondents included 2 men and 2 women, all of them intermediate fluent users of English with a clearly recognizable Czech accent. The fluency requirement was particularly important as the participants were instructed to read each sentence as one breath group without any interruptions or dysfluencies. The group of native speakers was comprised of 2 males and 2 females, all standard British speakers. This further material is referred to as the *focus condition*. The respondents were recorded individually in a sound treated booth at 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. The recording sessions were supervised by one of the authors. The native English data from the two conditions (radio × non-professionals) were examined separately to make sure that possible differences are not caused by different speech material or speaker status. (This implies that similar mean F0 values are desirable for both professional and non-professional samples.)

The utterances were split into breath groups (stretches of speech between two intakes of breath) and F0 tracks were extracted by the autocorrelation method in Praat (Boersma & Weenink, 2014) with a 10-ms time step between measuring points and subsequently turned into Praat PitchTier objects where the F0 contour was smoothed by a 10-Hz filter. The contours were then carefully inspected and manually corrected to diminish measurement errors such as spurious measurements in creaky phonation or voiceless portions of the signal, octave jumps, or failures to detect F0 in breathy voice of low amplitude. The final step before the measurement entailed the interpolation of voiceless portions in F0 contours in order to approximate the flow of speech melody to human perception (*cf.* Volín & Bartůňková, 2015). Previous research investigating the impact of interpolation on distributional measures suggests that in individual cases the F0 descriptors can be changed quite profoundly. However, in larger data sets, which our sample with more than 2300 F0 tracks definitely is, the changes are insignificant (*ibid.*).

In order to carry out the cross-linguistic comparison of speech melodies in native Czech, Czech-accented English and native English, long term distributional measures capturing the overall height of the speaker's voice (pitch level) and the range of frequencies covered by the speaker (pitch span) were calculated (Ladd, 1996), either in Hz or ST as appropriate. For level, the measures of central tendency were used: the mean and median, but also, in the same domain, we measured the F0 baseline. This metric is defined as the seventh percentile and proved to be a stable characteristic across various speaking styles and/or recording conditions (Lindh & Eriksson, 2007).

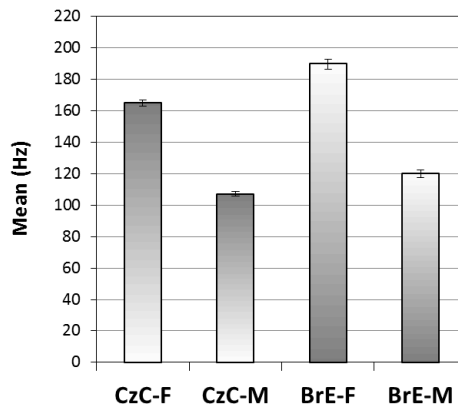
For span, the measures included the standard deviation, variation range (the difference between maximum and minimum), the 80-percentile range (i.e., the difference between the 90<sup>th</sup> and 10<sup>th</sup> percentile) and the quartile range (i.e., the difference between

the 75<sup>th</sup> and 25<sup>th</sup> percentile). The last two parameters display more stability than the maximum-minimum range, since certain portions of extreme values are cut off: 10% and 25% off each pole, respectively. The percentile values were computed in ST and normalized by each speaker's average. Finally, we also measured the gradient of the regression line in each breath group. This metric expresses in semitones per second the average rate of F0 declination and is computed by the least-sum-of-squares method. Statistical significance of the results was tested with two-way ANOVAs for independent measures with the respective F0 parameter as a dependent variable and with language and gender as independent variables. In Section 3.1 (Reference data) we provide two results: one for item analysis (each observation is a breath-group contour), the other for speaker analysis (each observation is a mean value of one of the speakers). In Section 3.2 (Focus data) speaker analysis is not possible as there are only eight respondents, so item analysis is a sole source of statistical significance quantification.

### 3. Results

#### 3.1 Reference condition

The pitch levels of the two investigated groups (read speech of native Czech and English professional speakers) show a significant difference between native Czech and native English speakers. The Czechs exhibit a significantly lower mean F0, which is statistically confirmed in analyses of variance:  $F(1, 1492) = 601.8$ ;  $p < 0.001$ , and  $F(1, 28) = 15.3$ ;  $p < 0.001$ . The effect can be observed both for male and female speakers. Mean values in Hz for the two languages and two genders are shown in Figure 1.



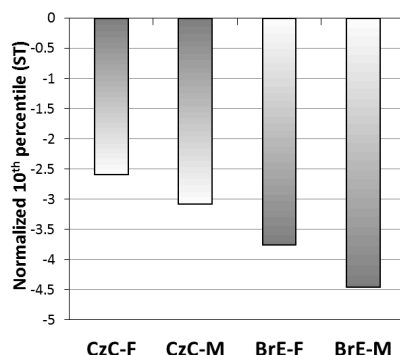
**Figure 1.** Mean F0 (in Hz) in the reference condition for two genders (M – male, F – female) and languages (CzC – Czech, BrE – English). Whiskers indicate 95% confidence interval.

Mean F0 of the Czech speakers is 165 Hz (female) and 107 Hz (male). The English speakers produced a mean F0 of 190 Hz (female) and 120 Hz (male). The significant difference between these two groups of speakers is also reflected by the two other indicators of pitch level, i.e., median and baseline value. The median values of the speakers are about 2-4 Hz lower than the mean. The baseline values are 141 Hz and 88 Hz for the female and male Czech speakers, respectively, and 150 Hz and 91 Hz for the female and male English speakers. The individual reference values are summarized below in Table 1.

Speaker group	Mean (Hz)	Median (Hz)	Baseline (Hz)	Minimum (ST)
CzC-F	165	162	141	4.2
CzC-M	107	105	88	-4.5
BrE-F	190	186	150	4.8
BrE-M	120	118	91	-4.3

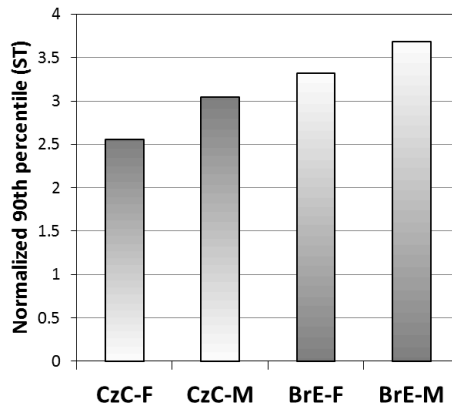
**Table 1.** Reference values for F0 level (in Hz and ST) for two genders (M – male, F – female) and two languages (CzC – Czech, BrE – English). *Baseline* refers to the 7<sup>th</sup> percentile (see Lindh & Eriksson, 2007). The raw minimum is related to 100 Hz.

Another measure, similar to the baseline value, is the 10<sup>th</sup> percentile (i.e., the low F0 indicator, disregarding 10 % of the lowest values). In Table 1 it is provided in its raw form. However, if it is normalized to the speaker's average, it can indicate individual distributional properties, namely symmetries or asymmetries in distribution of F0 values. Unlike the raw measure, which shows how low the speaker's voice is relative to an arbitrary anchor (100 Hz in our case), the normalized value reveals how far the speaker moves below his/her average. The results concerning the normalized 10<sup>th</sup> percentile values are shown in Figure 2. The native English speakers display significantly lower F0 values than the native Czech speakers:  $F(1, 1492) = 363.7$ ;  $p < 0.001$ , and  $F(1, 28) = 28.1$ ;  $p < 0.001$ , which means that they make more use of lower frequencies than the Czechs (cf. Table 1 for their higher central tendencies, though!).



**Figure 2.** Normalized average 10<sup>th</sup> percentile values of F0 (in ST) in the reference condition for two genders (M – male, F – female) and languages (CzC – Czech, BrE – English).

Czech female speakers have the highest 10<sup>th</sup> percentile value (-2.6 ST below their average), whereas the English male speakers the lowest (-4.5 ST below their average). Figure 3 shows the counterpart measure to the previous – the normalized 90<sup>th</sup> percentile values (i.e., an indicator of the high pitch boundary, disregarding 10% of the highest values) for the two investigated languages.



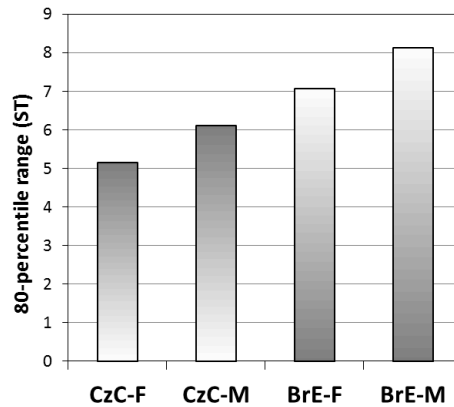
**Figure 3.** Normalized average 90<sup>th</sup> percentile values of F0 (in ST) in the reference condition for two genders (M – male, F – female) and languages (CzC – Czech, BrE – English).

The native English speakers increase their range to higher frequencies significantly more than the Czechs ( $F(1, 1492) = 56.8; p < 0.001$ , and  $F(1, 28) = 9.1; p < 0.01$ ). The English male speakers achieve a value as high as 3.7 ST above their average, whereas Czech female speakers only rise 2.6 ST above their average. Interestingly, the Czech speakers produce the same departure below and above their average, while the English speakers' deviations are slightly asymmetric – they go lower than higher (cf. Figure 2 vs. Figure 3).

In both languages, the difference between male and female speakers was very similar – the females stayed nearer to the average pitch level than the males and this effect was statistically significant ( $F(1, 1492) = 15.2; p < 0.001$  for the 90<sup>th</sup> percentile and  $F(1, 1492) = 62.3; p < 0.001$  for the 10<sup>th</sup> percentile).

The maximum and minimum F0 values of the reference speakers reflect the same effects and tendencies as the percentile values and, therefore, are not represented in graphs.

To summarize, the results of the last two analyses indicate that native English speakers exhibit larger deviations from their average pitch level than the native Czechs, and the deviations go in both directions, i.e., they use higher as well as lower frequencies than the Czech group. Furthermore, there is a difference between the male and female speakers – the females display smaller deviations from the average in both directions.

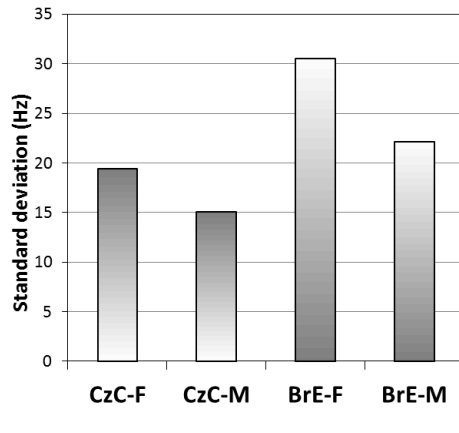


**Figure 4.** Mean values of the 80-percentile range of F0 (in ST) in the reference condition for two genders (M – male, F – female) and languages (CzC – Czech, BrE – English).

These findings are corroborated by the pitch span measurements. The 80-percentile range of F0 values is also significantly different across the two languages:  $F(1, 1492) = 301.7$ ;  $p < 0.001$ , and  $F(1, 28) = 18.9$ ;  $p < 0.001$ . The native English speakers display a greater F0 range (on average 7.1 ST for female and 8.1 ST for male speakers) than the native Czech speakers, whose percentile range is only 5.2 ST for female and 6.1 ST for male speakers. These differences are also reflected by other measurements of pitch span (minimum-maximum and quartile range). The difference between male and female speakers is also significant:  $F(1, 1492) = 61.2$ ;  $p < 0.001$ , and  $F(1, 28) = 5.01$ ;  $p < 0.05$ , the female speakers' pitch span being generally smaller. Figure 4 shows the mean values. The quartile range (25<sup>th</sup> to 75<sup>th</sup> percentile) shows very similar tendencies to the 80-percentile range.

Figure 5 below shows another measure of the variation in the data, the standard deviation (SD). This measure is provided here because it is widely used, and some researchers may wish to use our result for cross-linguistic comparisons. However, we are aware of the fact that SD should only be used for symmetrical distributions. Again, the effect of language is statistically significant:  $F(1, 1492) = 513.7$ ;  $p < 0.001$ , and  $F(1, 28) = 28.3$ ;  $p < 0.001$ , the Czech speakers display lower standard deviations than the English. In contrast to the 80-percentile range values, female speakers have higher SDs than their male counterparts. This is obviously an artefact of measurements in Hertz, since the female speakers operate in higher F0 bands and relative to human perception, Hertz is an exponential unit.





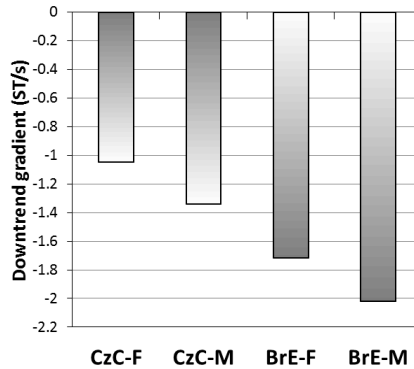
**Figure 5.** Standard deviation of F0 (in Hz) in the reference condition for two genders (M – male, F – female) and languages (CzC – Czech, BrE – English).

Table 2 summarizes the span measures for our reference sample. The ranges in semitones provide a consistent illustration, while standard deviation shows conflicting relationship between men and women due to the nature of the Hertz unit. At this point, we should also emphasise that despite its popularity, variation range is the least stable as it hinges on two extreme measures.

Speaker Group	Variation Range (ST)	80-perc. Range (ST)	Quartile Range (ST)	SD (Hz)
CzC-F	8.8	5.2	2.8	19.4
CzC-M	10.8	6.1	3.2	15.1
BrE-F	12.2	7.1	3.8	30.6
BrE-M	14.0	8.1	4.2	22.1

**Table 2.** Reference values of F0 span for two genders (M – male, F – female) and two languages (CzC – Czech, BrE – English). *SD* (standard deviation) is given in Hz, the *80-percentile range* (10<sup>th</sup> to 90<sup>th</sup> percentile), *quartile range* (25<sup>th</sup> to 75<sup>th</sup> percentile) and *variation range* (minimum to maximum) are given in ST.

Differences in the downtrend gradient were also measured through all-point regression analysis. The native Czech and native English speakers differ significantly:  $F(1, 1492) = 17.7$ ;  $p < 0.001$ , and  $F(1, 28) = 18.7$ ;  $p < 0.001$ , the English group displays generally a lower gradient value than the Czech group, which indicates a larger decline of F0 values throughout the utterance. The average gradient values are shown in Figure 6.

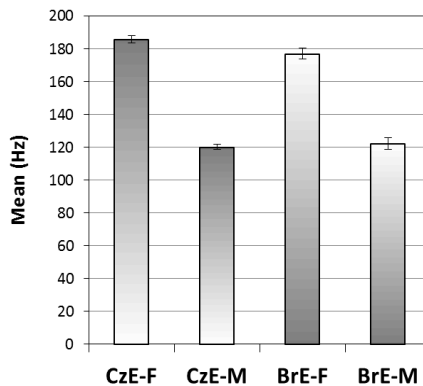


**Figure 6.** Average values of the downtrend gradient (in ST/s) in the reference condition for two genders (M – male, F – female) and languages (CzC – Czech, BrE – English).

The smallest gradient is found by the Czech female speakers, whose average decline is only about -1 ST/s. The Czech male speakers decrease their F0 by -1.3 ST/s, while the English speakers by -1.7 ST/s (female) and -2 ST/s (male). Even though the differences look similar, the two-way ANOVA confirms that while the effect of language is highly significant, the effect of gender is not ( $p > 0.1$ ) and, therefore, the male and female speakers in each group do not differ from each other above chance.

### 3.2 Focus condition

In this set of analyses, the groups of Czech and English non-professional speakers will be compared. Both groups read an identical set of English sentences in identical circumstances.

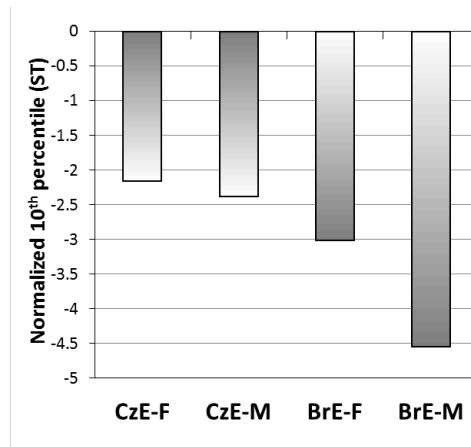


**Figure 7.** Mean F0 (in Hz) in the focus condition for two genders (M – male, F – female) and two languages (CzE – Czech English, BrE – English). Whiskers indicate 95% confidence interval.

The mean F0 values of Czech-accented English lie in between the reference English and Czech values, and the difference is statistically significant:  $F(1, 907) = 15.8$ ;  $p < 0.001$ . The English focus group displays very similar values, only the female mean is slightly lower (non-significantly) than for the Czech English focus group. The difference between the two groups of English speakers (reference and focus) is not significant ( $p > 0.05$ ). Figure 7 presents mean F0 values of the focus speakers.

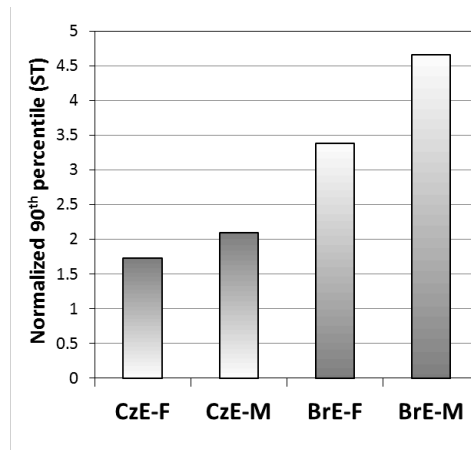
The Czech English female speakers have a mean F0 value of 186 Hz, the male speakers 120 Hz. The native English speakers produced a mean of 177 Hz (female) and 122 (male). The median values are almost the same as mean for the Czech groups, but 5 Hz lower for the English groups. The baseline values are 161 Hz and 102 Hz for the Czech English group, versus 146 Hz and 92 Hz for the native English group (females and males, respectively). Whereas the mean values of the male groups were very similar, their baseline values show a significant difference:  $F(1, 907) = 136.5$ ;  $p < 0.001$ .

The normalized 10<sup>th</sup> percentile values of the Czech-accented English are higher than for the native English focus group and both reference groups, which means that the pitch span of the Czech English focus speakers is the smallest also with respect to the lower frequencies. The difference is statistically significant:  $F(1, 907) = 390.7$ ;  $p < 0.001$ . Figure 8 shows the two focus groups (for reference groups see Figure 2). The difference between Czech English male and female group is not significant.



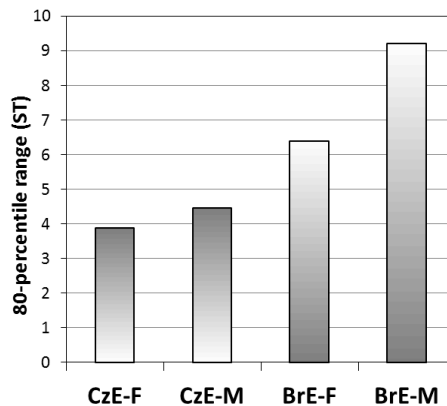
**Figure 8.** Normalized average 10<sup>th</sup> percentile values of F0 (in ST) in the focus condition for two genders (M – male, F – female) and two languages (CzE – Czech English, BrE – English).

Concerning the 90<sup>th</sup> percentile values, the results for the Czech English speakers show that the higher boundary of their pitch span is the lowest of all investigated groups. The Czech female focus speakers deviate only by 1.7 ST above their average, their male counterparts by 2.1 (see Figure 9). This is significantly lower than the Czech reference values and even more than both native English groups. The values of the English female focus group are very similar to the English female reference group, for males, the focus group has a higher 90<sup>th</sup> percentile values. Moreover, the English focus group show a very similar pattern to the 10<sup>th</sup> percentile values above, only the difference between males and females is smaller. Measurements of maxima concur with these findings.



**Figure 9.** Normalized average 90<sup>th</sup> percentile values of F0 (in ST) in the focus condition for two genders (M – male, F – female) and two languages (CzE – Czech English, BrE – English).

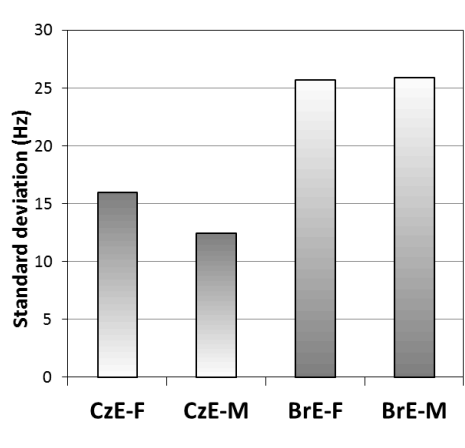
Pitch span indicators are represented in this section by the 80-percentile range (Figure 10) since the variation and quartile ranges behave in a very similar manner. The values produced by the Czech English speakers are lower than the reference values for Czech and English speakers, as well as for the English focus group. The Czech English group produced the pitch span of only 3.9 ST (female) and 4.5 (male speakers), while the English focus speakers reach 6.4 ST (female) and 9.2 ST (male speakers), which is lower than the female and higher than the male reference group, respectively.



**Figure 10.** Average values of the 80-percentile range of F0 (in ST) in the focus condition for two genders (M – male, F – female) and two languages (CzE – Czech English, BrE – English).

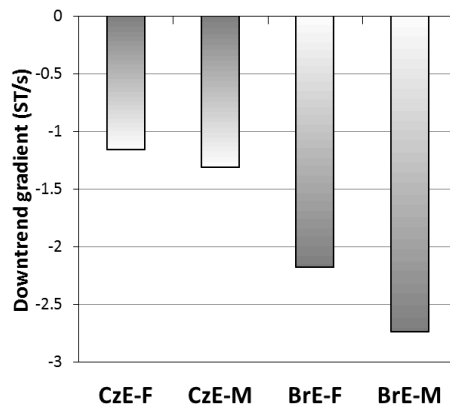
In general, there is a statistically significant difference between male and female speakers, the females having smaller pitch ranges than the males:  $F(1, 907) = 246.7$ ;  $p < 0.001$ . The quartile range confirms both effects of language and gender, and the resulting difference is even larger than with the 80-percentile range.

The standard deviations also show a highly significant effect of language  $F(1, 907) = 885.1$ ;  $p < 0.001$ , see Figure 11. However, Czech females show a higher standard deviation than Czech males, and the English group does not differ by gender at all. Similarly to the reference condition, this is the effect of measurements in Hertz.



**Figure 11.** Standard deviations of F0 (in Hz) in the focus condition for two genders (M – male, F – female) and two languages (CzE – Czech English, BrE – English).

Consequently, Czech-accented English displays the most reduced pitch spans of all the investigated groups. The deviations from the average pitch level are also smallest in both directions, above as well as below. Moreover, female speakers from both focus groups exhibit significantly smaller departures from their average, which result in a smaller pitch span than the male speakers.



**Figure 12.** Average values of the downtrend gradient (in ST/s) in the focus condition for two genders (M – male, F – female) and two languages (CzE – Czech English, BrE – English).

The downtrend gradients of the focus groups shows the same trend as in the reference groups (see Fig. 12, cf. Fig. 6). The effect of language is highly significant  $F(1, 907) = 313.3$ ;  $p < 0.001$ , while the effect of gender is only caused by the male native English focus speakers, who also exhibit the greatest gradient (almost -3 ST/s) indicating a relatively steep F0 decline throughout their breath groups.

The highest gradient value in this condition of -1.2 ST/s is the average of the female Czech English group, which is comparable with the reference speakers (-1 ST/s). The same holds true for the male Czech English speakers (-1.3 ST/s). The English focus groups display lower gradients than the reference English speakers. As in the previous analyses, the female speakers generally produced higher mean gradient values, which suggests that their F0 decline is not as steep as that of their male counterparts.

## 4. Discussion

Although distributional measures have been criticized for the lack of perceptual validity (Patterson, 2000), they seem to provide a logical point of departure for studies into the fundamental frequency domain of human speech. It has also been suggested over recent decades that their significance for diagnostic purposes needs to be acknowledged. One of the goals of our study was to provide reliable F0 measures obtained on a representative sample of speakers so that the indicators can be used for future reference in cross-linguistic comparisons. Our reference values portray Czech and English professional speakers; hence they guarantee a reasonable relationship to the population standards. Czech national radio and the BBC apparently select their newsreaders to fit the public expectations about the sound of their mother tongue.

The pitch level descriptors (Table 1 above), mean and median, found in the speech of English (British) men and women were higher than the same measures in the Czech sample. However, if we consider the baseline value and the raw minimum as alternative indicators of pitch level, the difference between native English and Czech speech decreases dramatically. This means that although the central tendency for BBC newsreaders is higher, they depart from it more than Czech radio speakers. The higher mean and median values could be caused by physiological or affective factors. Either the BBC officials offer the newsreading jobs to men and women with higher voices or it is expected from the British newsreaders to speak with greater involvement, which also leads to higher voice settings. The answer to this question cannot be based on our data, yet it is apparent that our randomly chosen non-professional speakers who read 112 independent sentences each do not mirror the trend, i.e., the non-professionals of British origin do not produce higher means and medians than their Czech counterparts.

Another very clear and statistically significant trend in our reference data was the greater pitch span in native English speech (Table 2 above). We captured the span in three range measures (variation range, 80-percentile range and quartile range) and standard deviation. All four metrics indicated smaller variation of F0 values in Czech speech. It is necessary to be cautious with the variation range since it depends on just two extreme measures (minimum and maximum) and can be quite volatile. It is also questionable whether human listeners base their perceptual assessments of the speaker's range on the two extremes in the contour. However, the ranges between the 10<sup>th</sup> and 90<sup>th</sup>

percentile, and between the 1<sup>st</sup> and 3<sup>rd</sup> quartile are relatively stable and together with the standard deviation and variation range they point in the direction indicated above.

Observing foreign-accented speech in our everyday life, we may notice clear influences of the mother tongue on sound patterns of the target language. Therefore, we might expect the span measures of Czech learners of English to fall somewhere between the native Czech and native English reference values. Interestingly, our data do not corroborate this generalization. The measures of span in Czech English were even lower than those in native Czech (compare Fig. 4 and Fig. 10). This outcome should not be attributed to the factor of speech style, since both British professionals and British non-professionals produced very similar values, and the Czech non-professionals read the same sentences under identical conditions. We could hypothesize that perhaps the uncertainty or even moderate anxiety associated with speaking a foreign language could enhance the tendency of Czech speakers to use narrower pitch ranges.

Our final acoustic measure – the downtrend gradient – is not easy to interpret. Generally, it is known that the overall trend to intonational declination is weaker in spontaneous speech than in read texts. Not much is known, however, about differences amongst languages. Our reference sample (32 speakers) produced a significantly steeper decrease of F0 values in the course of the utterances for British than for Czech speakers. The difference between men and women was unconvincing in this case. Interestingly, the difference between British newsreaders and non-professional speakers was significant. The gradual decline of F0 values is more moderate in newsreading than in isolated sentences. This trend, however, does not apply to Czech speakers. Both newsreaders and non-professional speakers produced very similar downtrend gradients. Again, measuring the impression on the listeners of this phenomenon poses an appealing future challenge.

The most inviting task for future research is to model and compare the actual shapes of the F0 contours on the level of the prosodic phrase, but also on the level of the phonological word. That has not been done up to date, yet it begs to be done. Quite importantly, our preliminary observations indicate that the Czech language with its stress fixed on the first syllable and with its prevalence of post-stress melodic rise (L\*+H) influences the shape of melodic contours in Czech-accented English in a thought-provoking manner.

### Acknowledgements

This study was supported by the grant provided by the Czech Science Foundation, GAČR, under no. 14-08084S.

## References

- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgements of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42, 529-555.
- Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Trouvain, J. (2014). Comparison of Pitch Range and Pitch Variation in Slavic and Germanic Languages. In N. Campbell, D. Gibbon, & Hirst, D. (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 776-780). Dublin: TCD.
- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer. Version 5.4.06, retrieved from <http://www.praat.org/>.
- Derwing, T., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48, 393-410.
- Dolson, M. (1994). The Pitch of Speech as a Function of Linguistic Community. *Music Perception*, 11(3), 321-331.
- Gilbert, J. (2014). Myth 4: Intonation is hard to teach. In J. Levis (Ed.), *Pronunciation Myths: Applying Second Language Research to Classroom Teaching*. Ann Arbor, MI: University of Michigan Press.
- Hahn, L. D. (2004). Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals. *TESOL Quarterly*, 38(2), 201-223.
- Holm, S. (2007). The Relative Contributions of Intonation and Duration to Intelligibility in Norwegian as a Second Language. *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences* (pp. 1653-1656). Saarbrücken: IPA.
- Kamiyama, T. (2004). Perception of Foreign Accentedness in L2 Prosody and Segments: L1 Japanese Speakers Learning L2 French. *Proceedings of Speech Prosody 2004*. Nara, Japan: ISCA.
- Keating, P., & Kuo, G. (2010). Comparison of speaking fundamental frequency in English and Mandarin. *UCLA Working Papers in Phonetics*, 108, 164-187. Los Angeles: University of California.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lindh, J., & Eriksson, A. (2007). Robustness of long-time measures of fundamental frequency. *Proceedings of Interspeech 2007* (pp. 2025-2028). Antwerp: ISCA.
- Mennen, I., Schaeffler, F., & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English Speakers. *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences*. Saarbrücken: IPA.
- Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice* (pp. 53-76). Berlin: Mouton de Gruyter.
- Munro, M. J., & Derwing, T. (2005). Second language accent and pronunciation teaching: A research based approach. *TESOL Quarterly*, 39(3), 379-397.
- Nolan, F. (2006). Intonation. In B. Aarts & A. McMahon (Eds.), *Handbook of English Linguistics*. Oxford: Blackwell.
- Patterson, D. (2000). A linguistic approach to pitch range modelling. Unpublished doctoral dissertation. University of Edinburgh.



- Reed, M., & Jones, T. (2015). The Melody of English: Research and Resources for Teaching the Pragmatic Functions of Intonation. IATEFL PronSIG webinar held 17<sup>th</sup> February 2015.
- Rogerson-Revell, P. (2011). *English Phonology and Pronunciation Teaching*. London: Continuum.
- Volín, J., & Bartůňková, H. (2015). Assets and Liabilities of Simple Descriptors of Fundamental Frequency Tracks. In O. Niebuhr & R. Skarnitzl (Eds.), *Tackling the Complexity in Speech*. In print.
- Wells, J. C. (2006). *English Intonation*. Cambridge: Cambridge University Press.
- Wichmann, A. (2005). The Role of Intonation in the Expression of Attitudinal Meaning. *English Language and Linguistics*, 9(2), 229-253.