PROSODIC PHRASING IN GOOD PUBLIC SPEAKERS IN CZECH AND ENGLISH

Hana Hledíková

3. ročník bakalářského studia

Anglistika-amerikanistika, Česká jazyk a literatura

Fonetický ústav FF UK

1. Introduction

At the suprasegmental level (above the level of individual phonemes), speech is characterized by features such as intonation, tempo, rhythm, etc. The term used to cover these characteristics is prosody. Prosody is not merely an additional decorative element, but carries linguistic meanings and has important functions, many of them pragmatic in nature. It plays an important role in helping the listener process what is being said and it is an important factor in forming the impression a speaker makes on the listener.

In this study, we are interested in differences between prosodic phrasing in good speakers of Czech and English. We investigate the melodic and temporal characteristics of prosodic phrasing, focusing on the stylistic domain of public speeches, and try to describe what is typical for this speaking style and how it differs in Czech and English.

1.1. Prosodic Phrasing

Prosodic phrases are units into which the flow of speech is divided, which helps to organize it by grouping certain words together and plays a significant role in the listener's comprehension of what is being said. (Frazier, Carlson & Clifton, 2006, s. 246). The prosodic phrase is a unit defined as "the domain of a perceptually coherent intonational contour" (Shattuck-Hufnagel & Turk, 1996, s. 210) that is delimited by prosodic boundaries. The strongest prominence of the prosodic phrase is the nuclear stress, which is realized on the tonic syllable. It tends to occur in a word near the end of the phrase, usually on the stressed syllable of the last content word (Féry, 2017, s. 61). This word carries the tone – an elementary melodic movement which carries a very general meaning. It is possible to place the tone on a different word to emphasize it.

To a certain degree, prosodic structure is related to the syntactic structure of an utterance (Shattuck-Hufnagel & Turk, 1996, s. 196). However, syntax and prosody are not isomorphic, and prosody is not fully determined by syntax. A sentence with a given syntactic structure can have more possible prosodic realizations, and predictions about prosodic boundary placements in an utterance can be made in terms of likelihood rather than givenness (Cutler, Oahan & Van Donselaar, 1997, s. 170).

1.2. Prosodic Boundaries

Prosodic phrases are separated by prosodic boundaries. They are usually signalled by melodic and temporal features (melodic movement, final lengthening), sometimes by a pause. In the ToBI system of transcription, there are five break indices to indicate the strength of

separation of each two words, 0 being the strongest degree of connection (meaning that the words are connected by connected speech processes), and 4 being the strongest degree of disconnection. Break indices 3 and 4 are defined as marking the end of higher level prosodic constituents – the intermediate and intonation phrase (Beckman & Elam, 1997, s. 31-32).

1.3. Prosody in Czech and English

Based on naïve observation, Czech intonation is generally quite flat compared to English. This is largely due to an overall narrower pitch range, which has been investigated by Volín et al. (2015), addressing "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" (Volín et al., 2015, s. 109). Pitch span of English and Czech professional news-readers was measured and found to be significantly lower in Czech than in English.

Czech speech in everyday communication also seems to be characterized by longer prosodic phrases. Dividing the utterance into a smaller number of longer prosodic phrases also contributes to the perceived flatness of intonation, because there are longer stretches of speech without stronger melodic movements.

In this study, we investigate the differences between Czech and English prosodic phrasing in good public speakers. We predict that the two languages differ in their prosodic phrasing (length of phrases and melodic variation within them) and that prosodic phrasing is influenced by stylistic factors, such as the communicative situation and the speaker's competence. Although the intonation of spontaneous Czech speech in everyday communication is characteristically flat, good public speakers may use different strategies in their prosody to make a good impression on the audience.

2. Method

2.1. Material

As material for the analysis, we have chosen TEDTalks in American English and Czech. 15 speakers in each language were selected based on subjectively perceived speaker quality from approximately 40 talks found on YouTube. We assume that the fact that the speaker has taken part in a TEDTalk conference is a certain guarantee of his or her competence in itself, but to ensure the speakers' quality, we have conducted a perception test in which participants were asked to evaluate the 15 selected speakers in each language. They were played a 30-second segment from their speeches and asked to express their willingness to employ the speaker as

their spokesperson on a 7-point scale. English and Czech speakers were evaluated separately, we have asked two different groups of 8 participants to evaluate the 15 recordings in each language. 10 speakers in each language who have received the best overall score were selected for the analysis. This choice of material enables us to compare English and Czech on fairly homogenous samples. The talks are given in similar conditions, they are comparable in length (all are between 15 and 20 minutes long) and they are produced by good public speakers.

2.2. Analysis

The recordings were divided into shorter segments (approximately 60 seconds long) and automatically segmented by means of Prague Labeller (in the case of Czech; Pollák, Volín & Skarnitzl, 2007) and P2FA (in the case of English; Yuan & Liberman, 2008) forced alignment, which yielded the approximate placement of phone boundaries. Approximately 5 minutes of speech from each speaker were analysed.

Prosodic phrasing was labelled manually in Praat. We used break indices 3 and 4, 4 for the strongest type of boundary which is clearly perceptible and usually has both a clear melodic movement and final lengthening or in some cases a pause, 3 for a weaker boundary with a smaller melodic movement. Additional marking of "p" is used "to convey some sort of prosodic disfluency – for example, an abrupt cutoff after a false start or a perceptible prolongation or pause which sounds as if the speaker were hesitating while searching for the next word" (Beckman & Elam, 1997: 32). We have also marked the word carrying nuclear stress in each phrase.

We used the annotated data to measure temporal and *f0* characteristics. The variables measured were:

- number of syllables per prosodic phrase
- number of words per prosodic phrase
- speech rate in syllables/second
- standard deviation (SD) of *f0* in each prosodic phrase in semitones (ST)
- SD of f0 in the nuclear part of the phrase (i.e., during realization of the tone) in ST
- Cumulative Slope Index (CSI) in each prosodic phrase in ST/syllable (Hruška & Bořil, 2017)

The number of words and syllables per prosodic phrase was extracted using a script in Praat. Words were counted directly from the automatically segmented word tier, syllables were counted as the number of vowels in English, and as the number of vowels and syllabic

consonants in Czech. Syllabic consonants in Czech were defined as [r] or [l] between two consonants. The cases where a word ends in a syllabic consonant and is followed by a word starting with a vowel were not counted (but the data includes only a very small number of these cases). There was no need to define syllabic consonants in English, because words where a syllabic consonant can appear, such as "intervention" or "electromagnetism," were always automatically transcribed with [ə] followed by a consonant. Speaking rate in syllables/second was measured by dividing the extracted number of syllables per prosodic phrase by phrase duration in seconds (without pauses).

Fundamental frequency was extracted using autocorrelation in Praat with the default settings, except for pitch ceiling – *f0* was extracted in the frequency range of 75–320 Hz for male speakers and 75–450 Hz for female speakers. The extracted Pitch objects were smoothed using a 10-Hz filter (to exclude very small *f0* movements which do not affect intonation) and interpolated (to create a continuous *f0* contour even in unvoiced segments, which reflects the way intonation is perceived more accurately). Finally, the Pitch objects were converted into PitchTier objects which were used to measure the SD of *f0* in ST in each phrase and during realization of the tone, and CSI in each phrase. Semitones are perceptual units – they reflect the way we hear pitch differences. Measuring in semitones allows us to compare intonation ranges between speakers with different pitch levels using SD (which would not be possible in hertz, as perceptually equal ranges would have differing ranges in hertz depending on their pitch level). CSI is a measurement of *f0* variation calculated as "the sum of absolute frequency differences between subsequent pitch points divided by the duration of the measured speech segment" (Hruška & Bořil, 2017: 37). In other words, unlike the standard deviation, CSI takes into account multiple melodic movements in a phrase. In this case, the duration of the measured speech segment (prosodic phrase) is the number of syllables per prosodic phrase.

We tested the significance of the effects of language and type of prosodic break on the measured variables. We used linear mixed-effects (LME) modelling to determine how significant the influence of language and type of prosodic break was on the temporal and *f0* characteristics while also taking into account possible individual differences between speakers. The analysis was conducted in R (R Core Team, 2017) using the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). LME is used to analyse the influence of fixed effects, i.e. the independent variables under our control, and random effects, i.e. other factors that are not under our control but may have influenced the measured dependent variables. In this case, the fixed effects were LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4), and the random effect was SPEAKER intercept (which accounts for the fact that individual speakers

may significantly differ from each other in the measured characteristics) and by-SPEAKER slope for the effect of TYPE OF PROSODIC BREAK (which accounts for the fact that individual speakers may significantly differ in their realization of each type of the prosodic break). The model assumes that the residuals are normally distributed and homoscedastic, so the residual plots were visually inspected for deviations from normality and homoscedasticity. The significance of individual effects or interactions was tested by comparing the full model to a reduced model with the given factor or interaction excluded. We conducted Tukey post-hoc tests using the R package *multcomp* (Hothorn, Bretz & Westfall, 2008) to test specific pairwise comparisons (e.g. the significance of the difference in speaking rate between BI3 in English and BI3 in Czech). Plots showing mean values of the measured variables and confidence intervals were created using the *effects* package (Fox, 2003).

3. Results

3.1. Temporal characteristics

3.1.1. Number of syllables per prosodic phrase

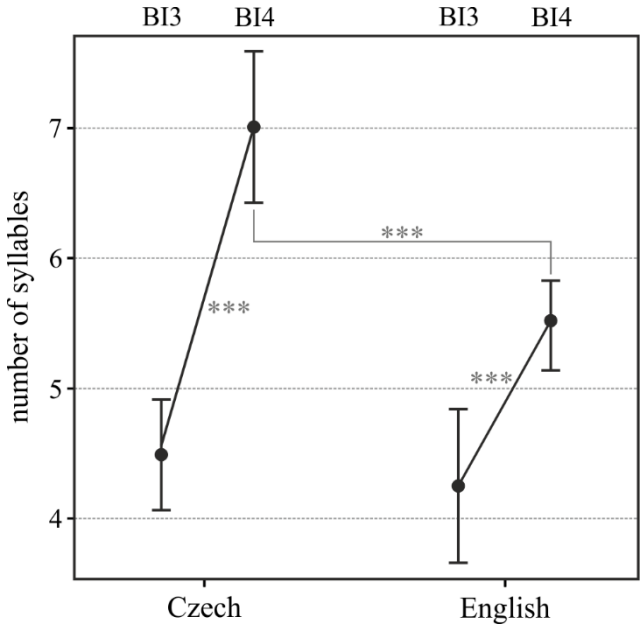

Figure 1. Number of syllables per prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4). The asterisks in this figure and all subsequent figures show statistical significance: *** $p < 0.001$, ** $p < 0.05$.

Mean values and confidence intervals are shown in **Figure 1.** Likelihood ratio tests comparing the full model (syllables ~ language + BI + (1+BI|speaker)) with the model without the effect in question show that both the effect of LANGUAGE and TYPE OF PROSODIC BREAK is significant. LANGUAGE significantly affected the number of syllables per prosodic phrase ($\chi^2(1) = 7.80$, $p < 0.01$): phrases in English are generally shorter by about 0.89 (± 0.25 standard

errors) syllables. TYPE OF PROSODIC BREAK significantly affected the number of syllables per prosodic phrase ($\chi^2(1) = 25.18$, $p < 0.0001$), phrases ending in a stronger prosodic break are generally longer by about 0.5–3.7 syllables (the differences in length vary between individual speakers).

The residuals show a certain degree of heteroscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is on the borderline of convergence, the interaction is significant: $\chi^2(1) = 6.98$, $p < 0.01$.

Tukey post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.001$), the difference between BI4 in Czech and BI4 in English is significant ($p < 0.001$) and the difference between BI3 in Czech and BI3 in English is not significant ($p > 0.9$).

Because some phrases are realized with a hesitation, usually a filled pause while the speaker is searching for the next word (labelled "p" in our data), we excluded those and also looked at the length of phrases without hesitations. **Table 1** shows that hesitations are much more frequent in phrases ending in BI3 (236 out of 573, that is almost a half, is realized with a hesitation), meaning that there is only a small number of phrases ending in BI3 without a hesitation left, which poses a problem for the analysis. Therefore, only phrases ending in BI4 without hesitations are considered.

language	BI	inter	n
Czech	3	no	136
Czech	4	no	1750
English	3	no	201
English	4	no	2133
Czech	3	yes	168
Czech	4	yes	110
English	3	yes	68
English	4	yes	38

Table 1. Number of prosodic phrases with and without hesitation ("inter")

The results are shown in **Figure 2**. LANGUAGE significantly affected the number of syllables per prosodic phrase ($\chi^2(1) = 15.60$, $p < 0.0001$), phrases in English are generally shorter by about 1.72 (± 0.37 standard errors) syllables. The residuals are normally distributed and slightly heteroscedastic. Compared to the difference between English and Czech in all prosodic phrases (i.e. also in those ending with BI3 and with hesitations), which was about 0.89 syllables, the difference here is almost twice as large. It seems that many of the shorter phrases ending in BI4 in Czech can be accounted for by an interruption due to the speaker's hesitation.

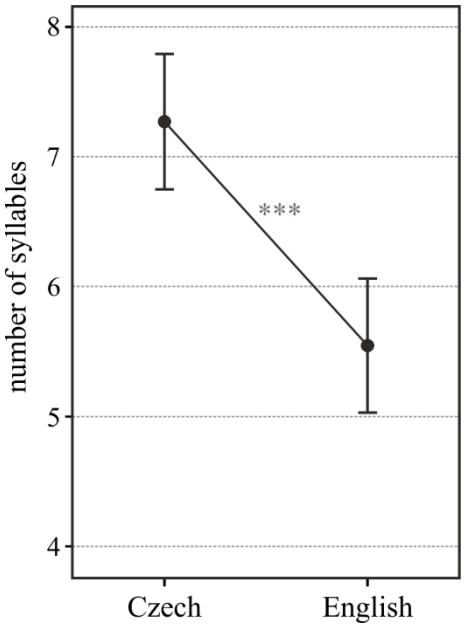

Figure 2. Number of syllables per prosodic phrase in BI4 without hesitation depending on LANGUAGE (Czech × English).

3.1.2. Number of words per prosodic phrase

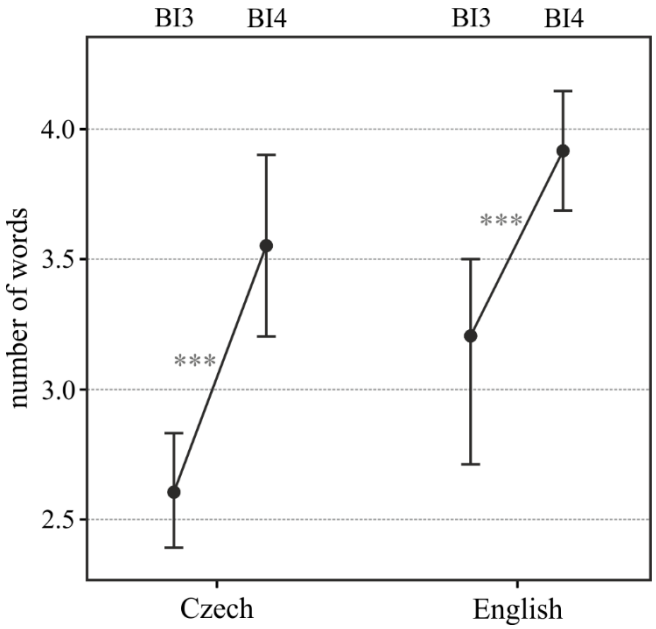

Figure 3. Number of syllables per prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 3**. The effect of LANGUAGE on the number of words per prosodic phrase is marginally significant ($\chi^2(1) = 2.88$, $p < 0.1$), phrases in English are generally longer by about 0.43 (± 0.13 standard errors) words. This is to be expected, because even though Czech phrases are generally longer in terms of syllables, English, which is an analytical language, uses many short words with a grammatical function

(such as articles), as opposed to Czech. TYPE OF PROSODIC BREAK significantly affected the number of words per prosodic phrase ($\chi^2(1) = 23.47$, $p < 0.0001$): phrases ending in a stronger prosodic break are generally longer by about 0.5–1.6 words (the differences in length vary between individual speakers, and there is one exceptional speaker whose phrases ending with BI3 are longer than those ending with BI4 by about 0.02 words).

The whole model manifests borderline singularity, the residuals show a show a satisfactory level of normality and homoscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is singular and could not be carried out.

Post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.0001$). The difference between BI3 in Czech and BI3 English and BI4 in Czech and BI4 in English is not significant ($p > 0.1$).

Figure 4 shows the results for phrases ending in BI4 and without hesitation. The effect of LANGUAGE on the number of words per prosodic phrase is not significant ($\chi^2(1) = 1.66$, $p > 0.1$). The residuals are slightly heteroscedastic.

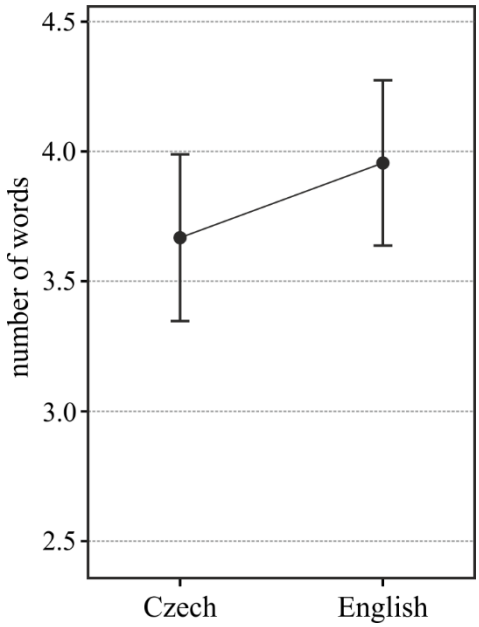

Figure 4. Number of words per prosodic phrase in BI4 without hesitation depending on LANGUAGE (Czech × English).

3.1.3. Speaking rate

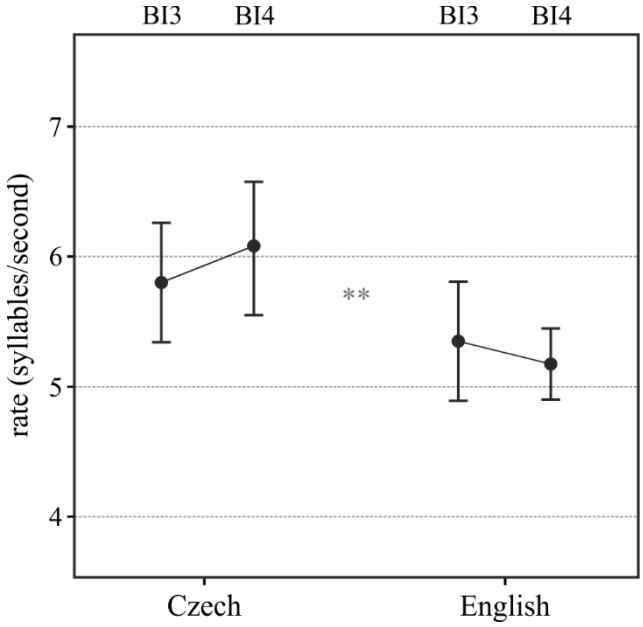

Figure 5. Speaking rate depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 5**. LANGUAGE significantly affected speaking rate ($\chi^2(1) = 5.38$, $p < 0.05$), speaking rate in English is generally lower by about 0.76 (± 0.29 standard errors) syllables/second. The effect of TYPE OF PROSODIC BREAK is not significant ($\chi^2(1) = 0.26$, $p > 0.6$), differences in rate in phrases ending with BI3 and BI4 are different in individual speakers, some having higher speaking rate in BI4, some lower.

The residuals are normally distributed and homoscedastic.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK failed to converge and could not be carried out.

3.2. *f0* characteristics

3.2.1. Standard deviation of *f0* in each prosodic phrase

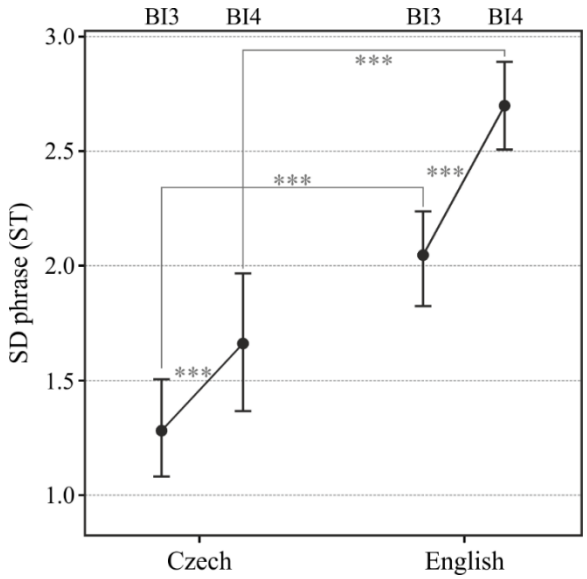

Figure 6. Standard deviation of *f0* in each phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 6**. LANGUAGE significantly affected the SD of *f0* in each prosodic phrase ($\chi^2(1) = 15.98$, $p < 0.0001$), SD in English is generally higher by about 0.78 (± 0.15 standard errors) ST. TYPE OF PROSODIC BREAK significantly affected the SD of *f0* in each prosodic phrase ($\chi^2(1) = 26.74$, $p < 0.0001$), the differences between SD in prosodic phrases ending with BI3 and BI4 range from a slight difference in some speakers (about 0.08 ST in the speaker with the lowest difference) to a considerably large difference (about 0.94 ST in the speaker with the highest difference).

The residuals show a certain degree of heteroscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is singular and could not be carried out.

Post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.0001$). The difference between BI4 in Czech and BI4 in English is significant ($p < 0.0001$) and the difference between BI3 in Czech and BI3 in English is also significant ($p < 0.001$).

3.2.2 Standard deviation of *f0* in the nuclear part of the phrase

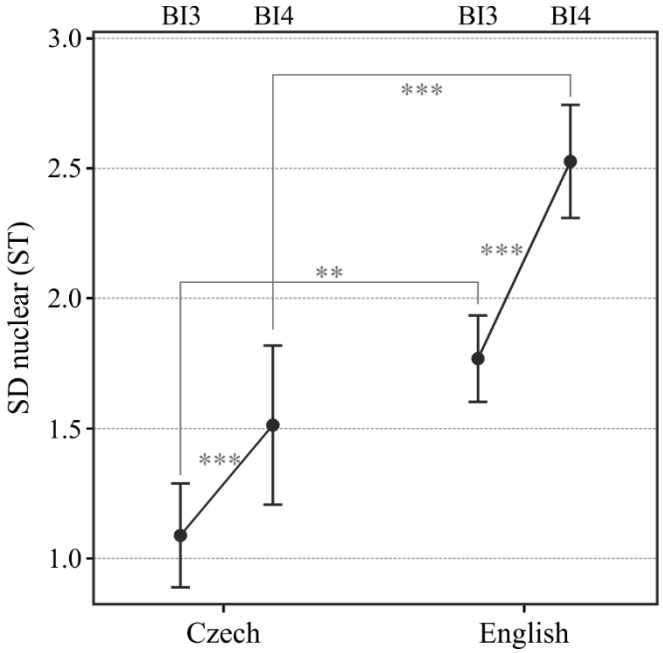

Figure 7. Standard deviation of *f0* in the nuclear part of the phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 7**. LANGUAGE significantly affected the SD of *f0* in the nuclear part of the phrase ($\chi^2(1) = 15.97$, $p < 0.0001$): SD in English is generally higher by about 0.67 (± 0.13 standard errors) ST. The model is on the borderline of convergence. TYPE OF PROSODIC BREAK significantly affected the SD of *f0* in the nuclear part of the phrase ($\chi^2(1) = 27.99$, $p < 0.0001$), the differences between SD in prosodic phrases ending with BI3 and BI4 range from a slight difference in some speakers (about 0.001 ST in the speaker with the lowest difference) to a considerably large difference (about 1.01 ST in the speaker with the highest difference).

The residuals show a satisfactory level of normality and homoscedasticity.

The test of interaction between LANGUAGE and TYPE OF PROSODIC BREAK is singular and could not be carried out.

Post-hoc tests show that the difference between BI3 and BI4 is significant both in Czech and in English ($p < 0.0001$). The difference between BI4 in Czech and BI4 in English is significant ($p < 0.0001$) and the difference between BI3 in Czech and BI3 in English is also significant ($p < 0.01$).

3.2.3 Cumulative Slope Index in each prosodic phrase

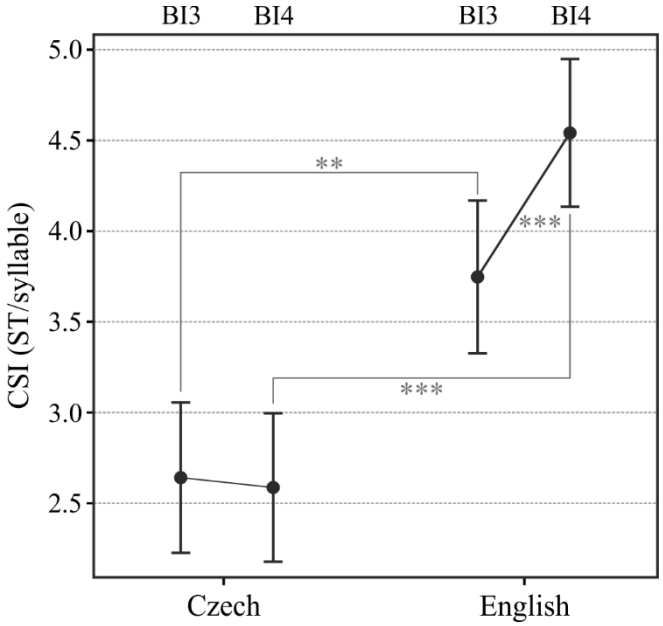

Figure 8. CSI in each prosodic phrase depending on LANGUAGE (Czech × English) and TYPE OF PROSODIC BREAK (BI3 × BI4).

Mean values and confidence intervals are shown in **Figure 8**. LANGUAGE significantly affected the SD of *f0* in the nuclear part of the phrase ($\chi^2(1) = 15.73$, $p < 0.0001$), CSI in English being higher by about 1.62 ST/syllable (± 0.27 standard errors). The test of the effect of TYPE OF PROSODIC BREAK is singular and could not be carried out. The difference between CSI in prosodic phrases ending with BI4 and BI3 is around 0 in Czech speakers, while English speakers tend to have higher CSI in prosodic phrases ending with BI4.

The residuals and show a certain degree of non-normality and heteroscedasticity.

The interaction between LANGUAGE and TYPE OF PROSODIC BREAK is significant ($\chi^2(1) = 9.35$, $p < 0.01$).

4. Discussion

Differences in prosody between Czech and English have been found both in temporal and *f0* characteristics. In accordance with expectations based on naïve observation of both languages in everyday communication, Czech was shown to have significantly longer prosodic phrases, narrower pitch range and less melodic variation than English. This may account for Czech intonation sounding more flat and monotonous in comparison to English. Czech speakers were also found to have a slightly higher speaking rate.

There were significant differences between prosodic phrases ending with a stronger (BI4) and weaker (BI3) prosodic boundary. Phrases ending with BI4 were longer in both syllables and words and exhibited higher pitch range both in the whole phrase and in the nuclear

part of the phrase. Melodic variability as measured by CSI was found to be higher in phrases ending with BI4 in English, but not in Czech.

We examined the speech of good public speakers. This means that the results relate to a specific stylistic domain and show features typical of public speeches. They cannot be easily generalized to the use of Czech and English in other contexts.

Even though Czech had longer phrases than English overall, most prosodic phrases were very short in both languages (mean length in English is 5.34 syllables, in Czech 6,62 syllables; phrases shorter than 10 syllables are notably more frequent than phrases longer than 10 syllables). If we exclude phrases realized with a hesitation, the relative amount of 1- and 2-syllable long phrases decreases, but the overall trend remains the same. This may be accounted for by the stylistic domain. It seems that good public speakers in both Czech and English divide their flow of speech more often by prosodic breaks to achieve a better effect on the audience. A more structured speech makes understanding the message easier for the listener, and shorter prosodic phrases with more melodic movement make the speech sound more emphatic and dynamic.

However, data from other communicative situations in both Czech and English would be needed to compare prosodic phrasing in good public speakers with prosodic phrasing in other speaking styles, such as spontaneous everyday communication.

References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015): *lme4: Linear Mixed-effects Models Using Eigen and S [R package version 1.1-8]*. Retrieved from <http://CRAN.R-project.org/package=lme4>.

Beckman, M. E., & Elam, G. A. (1997): *Guidelines for ToBI Labelling* (version 3.0). The Ohio State University Research Foundation.

Cutler, A., Oahan, D., & van Donselaar, W. (1997): Prosody in the Comprehension of Spoken Language: A Literature Review. *Language and Speech*, 40(2), s. 141-201.

Féry, C. (2017): *Intonation and Prosodic Structure*. Cambridge: Cambridge University Press.

Fox, J. (2003): Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8, s. 1–27.

Frazier, L., Carlson, K., & Clifton, C. Jr (2006): Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences,* 10(6), s. 244-249.

Hruška, R., & Bořil, T. (2017): Temporal Variability of Fundamental Frequency Contours. *Phonetica Pragensia – AUC Philologica*, 3, s. 35-44.

Hothorn, T., Bretz, F. & Westfall, P. (2008): Simultaneous inference in general parametric models. *Biometrical Journal*, 50, s. 346–363.

Pollák, P., Volín, J. & Skarnitzl, R. (2007): HMM-based phonetic segmentation in Praat environment. In: R. K. Potapova (ed.), *The 12th International Conference "Speech and Computer"*, SPECOM 2007, 15-18 October 2007, Moscow, Russia: Proceedings. Moscow: Moscow State Linguistic University, s. 537-541.

R Core Team (2017): *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

Shattuck-Hufnagel, S., & Turk, A. E. (1996): A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research*, *25*(2), s. 193-245.

Volín, J., Poesová, K., & Weingartová, L. (2015): Speech Melody Properties in English, Czech and Czech English: Reference and Interference. *Research in Language*, 13(1), s. 107-123.

Yuan, J., & Liberman, M. (2008): Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*. Retrieved from http://www.ling.upenn.edu/~jiahong/publications/c09.pdf