INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

The Acoustics of Word Stress in Czech As a Function of Speaking Style

Radek Skarnitzl[1], *Anders Eriksson*[2]

[1]Institute of Phonetics, Faculty of Arts, Charles University, Czech Republic
[2]Department of Linguistics, Stockholm University, Sweden
radek.skarnitzl@ff.cuni.cz, anders.eriksson@ling.su.se

Abstract

The study is part of a series of studies which examine the acoustic correlates of lexical stress in several typologically different languages, in three speech styles: spontaneous speech, phrase reading, and wordlist reading. This study focuses on Czech, a language with stress fixed on the first syllable of a prosodic word, with no contrastive function at the level of individual words. The acoustic parameters examined here are *F0-level*, *F0-variation*, *Duration*, *Sound Pressure Level*, and *Spectral Emphasis*. Values for over 6,000 vowels were analyzed.

Unlike the other languages examined so far, lexical stress in Czech is not manifested by clear prominence markings on the first, stressed syllable: the stressed syllable is neither higher, realized with greater F0 variation, longer; nor does it have a higher SPL or higher Spectral Emphasis. There are slight, but insignificant tendencies pointing to a delayed rise, that is, to higher values of some of the acoustic parameters on the second, post-stressed syllable. Since lexical stress does not serve a contrastive function in Czech, the absence of acoustic marking on the stressed syllable is not surprising.

Index Terms: lexical stress, speech prosody, Czech

1. Introduction

The present study is the latest addition to a series of studies describing the acoustics of lexical stress in a number of typologically different languages. An important point here is that the recorded material should have the same structure in all studies and that the analyses applied should be the same. This way, the results for the different languages may be immediately comparable. We have previously recorded databases for Brazilian Portuguese, English, Estonian, French, German, Italian, Swedish, and now Czech, and published results for Brazilian Portuguese [1, 2], Estonian [3], English [4], German [5], Italian [6], and Swedish [7, 8].

We have found that the acoustic correlates of stress are influenced by speaking style. Word list reading usually produces the most prototypical stress realization, typically described in pronunciation lexica. In spontaneous speech, acoustic correlates are often reduced and phrase reading falls somewhere in between. The speaking styles investigated in the studies are therefore *Wordlist Reading*, *Phrase Reading*, and *Spontaneous Speech*.

The study of the acoustics of lexical stress goes a long way back. Classical studies, such as those by Fry in the fifties [9], found that *F0*-level and variation, vowel duration and vowel amplitude correlated with word stress in English but not all to the same degree. These findings have been confirmed in a broad sense in studies of other languages like Polish [10], French [11], Swedish [12], and Spanish [13, 14]. Results from our previous studies are also compatible in a broad sense with these earlier studies. Amplitude, however, has not turned out to correlate very well with stress level or perception results but *Spectral Emphasis*, a related measure found to correlate with vocal effort [15] has been shown to correlate with stress in studies of Dutch [16, 17]. It has also been shown to play a role in American English [18, 19], British English [4], Italian [6] and Swedish [7, 8, 20].

The Czech language belongs to languages with fixed stress, placed on the first syllable of the prosodic word. Its function is thus demarcative, facilitating the segmentation of the stream of speech into stress groups. Lexical stress in Czech is independent of vowel quality, length, and morphological structure of the word: in the noun *posouvání* ("shifting", [ˈposo͡uvaːɲiː]), for instance, the first syllable will be stressed although it is a prefix, and it contains the only short vowel in the word. Monosyllabic words tend to behave like clitics, typically joined prosodically to the previous word.

Early studies of lexical stress in Czech, conducted by Janota, were based perceptually. In [21], the authors exploited the fact that in the sequence [sese], both syllables may potentially be stressed in connected speech because the word *se* has more meanings, and can be part of a longer word. When listeners were to identify the stressed syllable in the sequences excised from context, they used the acoustic properties, especially higher F0, to make their judgement. In delexicalized speech, however, listeners often identified as stressed syllables those which had a lower F0, SPL, as well as duration than those which were judged to be unstressed. Already in the 1970s, the authors proposed that Czech word stress is not determined by the acoustic prominence of the stressed syllable but by the acoustic configuration of the whole prosodic word.

More recent studies have focused especially on the configuration of F0 throughout stress groups and across their boundaries. It was found [22] that, in trisyllabic and longer stress groups, F0 tends to be higher on the second, post-stressed syllable than on the first, stressed syllable; the neutral configuration thus seems to be a delayed rise (L*+H) [23]. Other studies addressed spectral emphasis, but only examined trisyllabic pseudowords such as [tototo], with stress placed alternately on all three syllables; in such material, the stressed syllable was, indeed, realized with flatter tilt than the unstressed syllables [24], [25].

The current study is the first in which all principal acoustic correlates of lexical stress in Czech will be investigated at the same time (i.e., *F0-level*, *F0-variation*, *Duration*, *SPL*, and *Spectral Emphasis*) using the same type of speech material.

  http://dx.doi.org/10.21437/Interspeech.2017-417

2. Method

2.1. Speech material and speakers

In order to avoid the confounds of segmental-level variability, the analysis is based on identical speech material across all speaking styles. First of all, each speaker was recorded in a semispontaneous interview. They were free to choose the topic of the conversation. The interviews lasted 40–45 minutes. After the recordings were orthographically transcribed, we selected at least 60 target words which contained between three and six syllables; care was taken for the target words to be embedded in fluent context, not to appear in phrase-final or -initial contexts or carry focal accent. Based on these words, we prepared two manuscripts, one only with the words in isolation, the other containing the surrounding context. The wordlists and phrases were recorded approximately four weeks after the spontaneous session.

12 speakers (6 females, 6 males, 20–32 years) of Common Czech were recorded in a quiet room using a high-quality portable recorder, Edirol R-09HR. The original 48 kHz/16 bit wav recordings were later downsampled to 16 kHz for the acoustic analyses. A total of 6,337 vowels in 1,661 words were analyzed.

2.2. Acoustic parameters and analyses

As mentioned above, the aim of this study was to analyze all the main acoustic correlates of prominence. For every vowel, these were:

- *Fundamental frequency level*, defined here as the F0 median in the vowel, measured in semitones relative to 1 Hz; the median value was chosen so as to minimize the effect of outlier values.
- *F0 variation*, measured as the standard deviation of F0 in semitones; semitones were chosen so that variation is comparable across both female and male speakers.
- *Duration*, measured in milliseconds, expressed in this study as Log2(ms) in order to obtain a more normal distribution.
- *Sound pressure level* in decibels.
- *Spectral Emphasis*, in its simplified version used here, is computed as the difference between the energy of the entire spectrum (SPL_{full}) and the energy of the low-pass filtered segment using a cutoff frequency of 1.5 * $F0_{mean}$ (SPL_0); see [26]: *SE* (dB) = $SPL_{full} - SPL_0$.

Log-scales are thus used for all the examined parameters. The values of the acoustic parameters were extracted using a *Praat* [27] script.

All subsequent statistical analyses and data visualisation were performed in *R* [28], using the packages *effects* [29] and *ggplot2* [30].

The studies mentioned in the Introduction, especially those which analyzed F0 level [22], [23], as well as informal observations suggest that the second, post-stressed syllable in a stress group may behave differently in Czech from other unstressed syllables (i.e., third, fourth etc.). For that reason, we will recognize three levels of the STRESS factor: stressed, post-stressed, and unstressed. Apart from the position in the stress group, the acoustic manifestation of the vowels will be examined in relation to the factor STYLE (word, phrase, spontaneous; see section 2.1) and with some analyses SEX.

3. Results

3.1. Fundamental frequency level

As we can see in Figure 1, there are visible differences in F0 levels depending on both the STRESS factor (i.e., position of the syllable in the stress group) and the STYLE factor. Interestingly, males and females also differ in the relative pitch movements, particularly in isolated words: while female speakers seem to pronounce individual words in isolation with a falling melodic movement (with the drop between the post-stressed syllable and the remaining syllables being highly significant, according to Tukey HSD test: $p < 0.0001$), male speakers seem to prefer a rising–falling movement (the rise from the first to the second syllable is marginally significant, $p = 0.056$, and the subsequent fall to the remaining unstressed syllables is significant, $p < 0.01$).

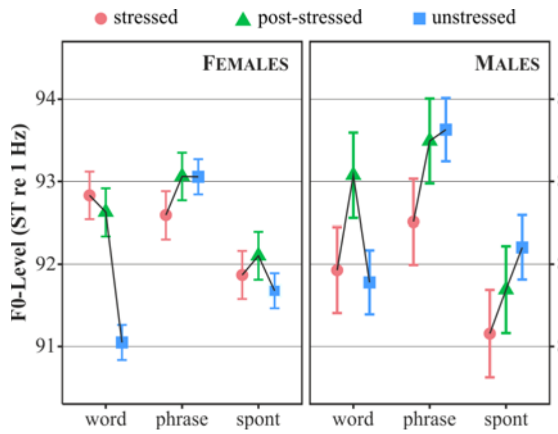

Figure 1: *Fundamental frequency level as a function of speaking style and position in the stress group.*

In the remaining speaking styles, phrases and spontaneous speech, we can observe a slight tendency to the above-mentioned delayed rise, in which the post-stressed syllable lies higher than the stressed one, but the difference between them is not significant ($p > 0.1$).

To consider the dataset as a whole, a Univariate ANOVA using F0 values in semitones as the dependent variable and STRESS and STYLE as fixed factors showed, for both sexes, significant main effects of STRESS [females: $F(2, 3394) = 23.9$; $p < 0.001$; males: $F(2, 2828) = 9.4$; $p < 0.001$] and STYLE [females: $F(2, 3394) = 65.2$; $p < 0.001$; males: $F(2, 2828) = 33.2$; $p < 0.001$], as well as a significant interaction between STRESS and STYLE [females: $F(4, 3394) = 22.6$; $p < 0.001$; males: $F(4, 2828) = 4.9$; $p < 0.001$].

The results presented so far do not confirm the presence of a delayed rise in F0, where the F0 would lie higher on the post-stressed syllable than on the stressed one. To investigate this issue a bit further, it is beneficial to consider the F0-level difference between the stressed and post-stressed syllable separately in each word. Figure 2 shows a histogram of the differences between F0-level in the second and first syllable; numbers higher than 0 indicate a rise on the second syllable relative to the stressed one, numbers lower than 0 a fall. Delayed rise (i.e., a prevalence of positive values) is more obvious in male speakers (mean F0 step is 0.94 ST, which is beyond the just noticeable difference [31]) than in female speakers (mean F0 step is 0.16 ST).

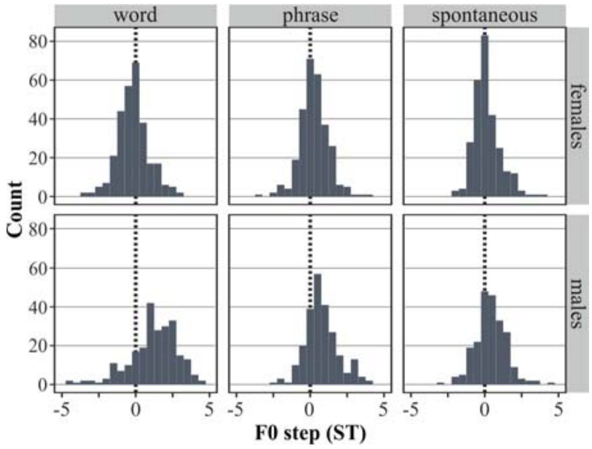

Figure 2: *Difference in F0-level between the post-stressed and stressed syllable (see text).*

3.2. Fundamental frequency variation

The results of a Univariate ANOVA with the standard deviation of F0 as the dependent variable show significant main effects of Stress [$F(2, 6101) = 21.1$; $p < 0.001$] and Style [$F(2, 6101) = 13.2$; $p < 0.001$] but no significant effect of Sex. In addition, there were significant interactions between Stress and Style [$F(4, 6101) = 12.6$; $p < 0.001$], and Stress and Sex [$F(2, 6101) = 6.0$; $p < 0.05$]. On closer examination (see Figure 3), however, no clear picture emerges: the vowel in the stressed, nor in the post-stressed syllable manifests a significantly higher F0 variability (Tukey HSD: $p > 0.1$). Greater variation can only be observed in female speakers' rendition of isolated words, where vowels in the last syllables are pronounced with a clear melodic change.

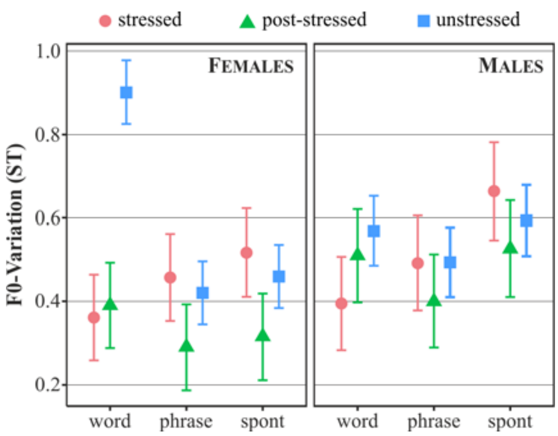

Figure 3: *Standard deviation of fundamental frequency as a function of speaking style and position in the stress group.*

3.3. Duration

Duration results, shown in Figure 4, are perhaps the most interesting presented thus far. A Univariate ANOVA shows significant main effects of Stress [$F(2, 6319) = 273.1$; $p < 0.001$], Style [$F(2, 6319) = 524.5$; $p < 0.001$], as well as Sex [$F(1, 6319) = 74.7$; $p < 0.001$]. In addition, there was a significant interaction between Stress and Style [$F(4, 6319) = 51.7$; $p < 0.001$]; the interaction between Stress and Sex is not significant.

There is a clear tendency for the vowel in the stressed syllable to be shortest and the vowels in subsequent syllables to be progressively longer. The difference in the duration between the stressed and post-stressed vowel is significant only in male speakers' rendition of isolated words (Tukey HSD: $p < 0.001$) and marginally significant in isolated words pronounced by female speakers ($p = 0.09$).

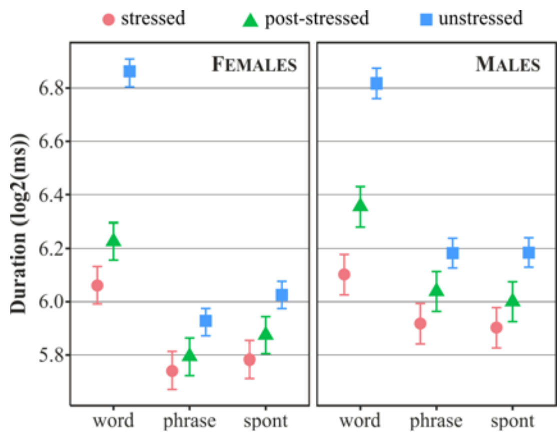

Figure 4: *Vowel duration (expressed in logarithmic form) as a function of speaking style and position in the stress group.*

The lengthening of the vowels in what we treat as unstressed syllables (i.e., third to sixth in the word) is intriguing, since the words originated from phrase-medial positions. The difference in the duration between the post-stressed and (the remaining) unstressed vowels was significant in isolated words ($p < 0.001$ for both males and females) and in male spontaneous speech ($p < 0.05$), with female spontaneous speech reaching only marginal significance ($p = 0.06$).

Figure 5 shows the vowel durations in more detail: the "unstressed" group, which comprised vowels in the third to sixth syllable in Figure 4, is now treated separately (six-syllable words are not shown for the sake of simplicity, as there were too few of them in the dataset). The progressive lengthening of vowels is, again, most evident in isolated words, where final deceleration is to be expected, but there is a clear tendency in the phrase reading and in spontaneous speech as well.

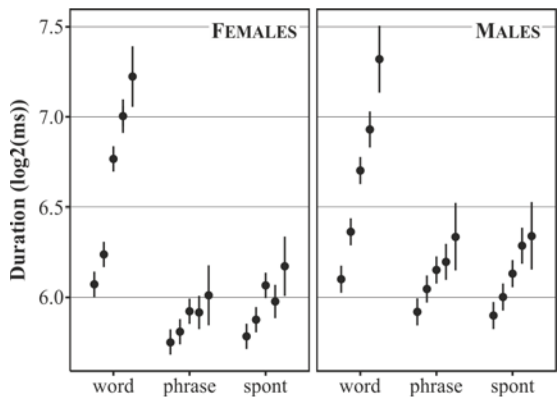

Figure 5: *Vowel duration (expressed in logarithmic form) as a function of speaking style and position in the stress group, for vowels in the first five syllables from left to right (see text).*

3.4. Sound pressure level

As we can see in Figure 6, SPL does not change significantly between the stressed and the post-stressed vowel in any of the speech styles (Tukey HSD tests: $p > 0.1$). However, there are clear main effects of STRESS [$F(2, 6319) = 56.0$; $p < 0.001$], STYLE [$F(2, 6319) = 365.4$; $p < 0.001$] and SEX [$F(1, 6319) = 86.9$; $p < 0.001$], as well as significant interactions between STRESS and STYLE [$F(4, 6319) = 20.5$; $p < 0.001$] and STRESS and SEX [$F(2, 6319) = 5.1$; $p < 0.01$]; all these seem to be due to the behaviour of the "unstressed" group.

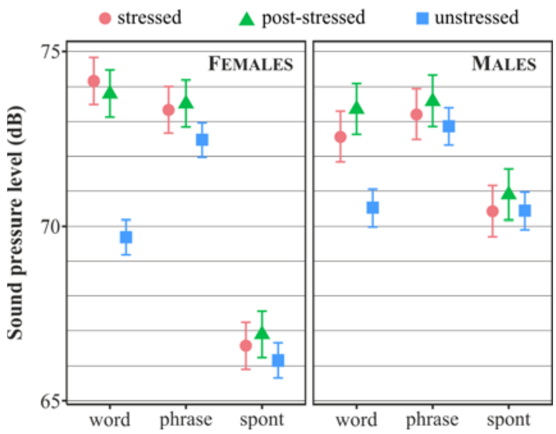

Figure 6: *Sound pressure level as a function of speaking style and position in the stress group.*

3.5. Spectral Emphasis

The results for Spectral Emphasis are shown in Figure 7. Given the importance of this parameter in other languages, it may be surprising that the Univariate ANOVA reveals the smallest number of significant main effects for Spectral Emphasis: STRESS [$F(2, 6319) = 10.0$; $p < 0.001$] and, not surprisingly, SEX [$F(1, 6319) = 382.3$; $p < 0.001$], given the generally breathier voice quality in female speakers [32]. The effect of STYLE was not significant. The only significant interaction which concerned the STRESS factor was between STRESS and STYLE [$F(4, 6319) = 2.9$; $p < 0.05$]. The data for males show a slight tendency for the post-stressed vowel to be more prominent (with Spectral Emphasis higher by about 0.5 dB), but the difference does not reach statistical significance according to the Tukey HSD post-hoc test ($p > 0.1$).

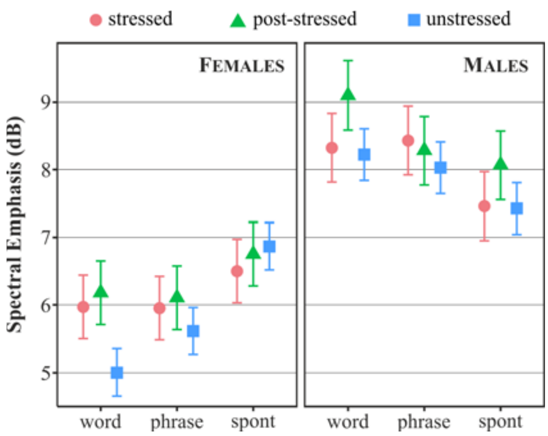

Figure 7: *Spectral Emphasis as a function of speaking style and position in the stress group.*

4. Discussion and conclusions

In comparison with previous studies in this series, which have examined the acoustics of stress in a number of languages, Czech clearly behaves in a different way. Being a language in which stress is fixed on the first syllable of a word and is not contrastive, it is to be expected that the acoustic marking of lexical stress – in the sense of higher values of the examined acoustic parameters on the stressed syllable – will be less salient, or even absent (*cf.* [33]). That is, indeed, what our data show: in none of the dimensions of prominence does the Czech stressed syllable manifest significantly higher values.

Table 1 compares the F-statistics of the examined parameters with the factor STRESS, i.e., the position of the vowel in the stress group. The F-statistics are given for spontaneous speech only. As we can see, *duration* is affected by the position of the vowel in the stress group the most, but not in the way one might expect: the duration of the stressed vowel is, in fact, shorter than the following vowels (*cf.* Figure 4), although the difference was not significant. Vowels which appear later in the word tend to be longer (Figure 5).

Table 1: *F-statistics of acoustic parameters with the STRESS factor, for spontaneous speech.*

Parameter	F-statistic
Duration	39.4
F0-Level	27.2
F0-Variance	4.3
SPL	1.5
Spectral Emphasis	1.0

The second most significant parameter, *F0-Level*, also behaves in a counter-intuitive way (if considered from a traditional perspective). The post-stressed (second) vowel lies slightly higher than the stressed one in both male and female speakers (*cf.* Figure 1), although the overall difference is not statistically significant for either sex. If we stipulate a strict *jnd* of 1 ST, then there is no movement in F0-Level (i.e., it falls within the <-1; 1> ST range) in 74% of the words in spontaneous speech, there is a post-stress rise in 20% of the words and a fall only in 6% (*cf.* Figure 2). In other words, a post-stress rise is approximately three times as frequent in our spontaneous data than a fall, but less frequent than reported in the studies mentioned above [22], [23]. As we can see in Table 1, the remaining parameters – *F0-Variance*, *SPL* and *Spectral Emphasis* – differ very little with respect to the position of the vowel in the stress group.

To conclude, lexical stress in Czech does not appear to be realized in a way that the stressed syllable would bear traditional marks of prominence: it tends to be shorter and slightly lower than the post-stressed syllable, with comparable SPL and Spectral Emphasis. The issue of configurations of acoustic parameters (mainly F0) throughout the stress group (see [21], [22]) has not been unequivocally resolved.

5. Acknowledgements

The research programme has been funded by the Swedish Research Council (VR) project *A typology for word stress and speech rhythm based on acoustic and perceptual considerations*, under grant 2007-2301. The first author was supported by the Charles University project Progres 4, *Language in the shiftings of time, space, and culture*.

6. References

[1] Barbosa, P. A., Eriksson, A., and Åkesson, J., "Crosslinguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese," in *Proc. Nordic Prosody 2012,* Frankfurt am Main: Peter Lang, 2013, pp. 97–106.

[2] Barbosa, P. A., Eriksson, A., and Åkesson, J., "On the robustness of some acoustic parameters for signaling word stress across styles in Brazilian Portuguese," in *Proc.Interspeech 2013,* Lyon, 2013, pp. 282–286.

[3] Lippus, P., Asu, E. L., and Kalvik, M.-L., "An acoustic study of Estonian word stress," in *Proc. Speech Prosody 2014,* Dublin, 2014, pp. 232–235.

[4] Eriksson, A. and Heldner, M., "The acoustics of word stress in English as a function of stress level and speaking style," in *Proc. Interspeech 2015,* Dresden, 2015, pp. 41–45.

[5] Behrens, J., "Die Prosodie des Wortakzentes in Abhängigkeit von Akzentlevel und Sprechstil," BA Thesis, Philosophischen Fakultät, Christian-Albrechts-Universität zu Kiel, 2013.

[6] Eriksson, A., Bertinetto, P.M., Heldner, M., Nodari, R., and Lenoci, G. "The Acoustics of Lexical Stress in Italian as a Function of Stress Level and Speaking Style", in *Proc. Interspeech 2016,* San Francisco, 2016, pp. 1059–1063.

[7] Eriksson, A., Barbosa, P. A., and Åkesson, J., "Word stress in Swedish as a function of stress level, word accent and speaking style," in *Proc. Nordic Prosody 2012,* Frankfurt am Main: Peter Lang, 2013, pp. 127–136.

[8] Eriksson, A., Barbosa, P. A., and Åkesson, J., "The acoustics of word stress in Swedish: A function of stress level, speaking style and word accent," in *Proc. Interspeech 2013,* Lyon, 2013, pp. 778–782.

[9] Fry, D. B., "Duration and intensity as physical correlates of linguistic stress," *Journal of the Acoustical Society of America*, vol. 27, pp. 765–768, 1955.

[10] Jassem, W., Morton, J., and Steffen-Bartóg, M., "The perception of stress in synthetic speech-lke stimuli by Polish listeners," *Speech Analysis and Synthesis*, vol. 1, pp. 289–308, 1968.

[11] Benguerel, A. P., "Physiological correlates of stress in French (Correlats physiologiques de l'accent en francais),"*Phonetica*, vol. 27, pp. 21–35, 1973.

[12] Fant, G. and Kruckenberg, A., "Notes on stress and word accent in Swedish," in *STL/QPSR,* vol. 2–3, 1994, pp. 125–144.

[13] Díaz-Campos, M., "The Phonetic Manifestation of Secondary Stress in Spanish," in *Hispanic Linguistics a the Turn of the Millennium,* Somerville, MA: Cascadilla, 2000, pp. 49–65.

[14] Vargas-Calderon, R., "Acoustic Analysis of Stress in the Spanish Spoken in Costa Rica (Analyse acoustique de l'accent de l'espagnol parle au Costa Rica)," *Travaux de l'Institut de Phonetique de Strasbourg*, vol. 18, pp. 1–23, 1986.

[15] Traunmüller, H. and Eriksson, A. "Acoustic effects of variation in vocal effort by men, women and children," *Journal of the Acoustical Society of America*, vol. 107, pp. 3438–3451, 2000.

[16] Sluijter, A. M. C., *Phonetic Correlates of Stress and Accent*. The Hague: Holland Academic Graphics, 1995.

[17] Sluijter, A. M. C. and van Heuven, V. J., "Spectral balance as an acoustic correlate of linguistic stress," *Journal of the Acoustical Society of America*, vol. 100, pp. 2471–2485, 1996.

[18] Campbell, N. and Beckman, M. E., "Stress, prominence, and spectral tilt," in *Intonation: Theory, Models, and Applications* Athens, 1997, pp. 67–70.

[19] Campbell, N. and Beckman, M. E., "Accent, stress, and spectral tilt," *Journal of the Acoustical Society of America*, vol. 101, p. 3195, 1997.

[20] Heldner, M., "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," *Journal of Phonetics*, vol. 31, pp. 39–62, 2003.

[21] Janota, P. and Palková, Z., "The auditory evaluation of stress under the influence of context," *Acta Universitatis Carolinae – Philologica, Phonetica Pragensia IV*, pp. 29–59, 1974.

[22] Palková, Z. and Volín, J., "The role of F0 contours in determining foot boundaries in Czech", in *Proc. 15th ICPhS*, Glasgow, 2003, pp. 1783–1786.

[23] Volín, J., "Z intonace čtených zpravodajství: výška první slabiky v taktu," *Čeština doma a ve světě*, 1-2/2008, pp. 89–96.

[24] Volín, J. and Zimmermann, J., "Spectral slope parameters and detection of word stress," in *Proc. Technical Computing Prague*, 2011, pp. 125–130.

[25] Weingartová, L. and Volín, J., "Short-term spectral slope measures and their sensitivity to speaker, vowel identity and prominence," *Akustické listy*, vol. 20, pp. 5–12, 2014.

[26] Traunmüller, H. and Eriksson, A., "Acoustic effects of variation in vocal effort by men, women, and children," *Journal of the Acoustical Society of America*, vol. 107, pp. 3438–3451, 2000.

[27] Boersma, P. and Weenink, D., *Praat: Doing Phonetics by Computer* (Version 5.4). Retrieved on October 14, 2014, from http://www.praat.org.

[28] R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org, 2016.

[29] Fox, J., "Effect displays in R for generalised linear models," *Journal of Statistical Software*, vol. 8, no. 15, pp. 1–27, 2003.

[30] Wickham, H., *ggplot2: Elegant graphics for data analysis (use R!)*. New York: Springer, 2009.

[31] Klatt, D. H., "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of speech perception," *Journal of the Acoustical Society of America*, vol. 53, pp. 8–16, 1973.

[32] Titze, I. R., "Physiological and acoustic differences between male and female voices," *Journal of Acoustical Society of America*, vol. 85, pp. 1699–1707, 1989.

[33] Cutler, A., Lexical stress. in *The Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds. Oxford: Blackwell Publishing, 2005, pp. 264–289.