The Acoustics of Word Stress in English as a Function of Stress Level and Speaking Style

Anders Eriksson, Mattias Heldner

Department of Linguistics, Stockholm University, Sweden
anders.eriksson@ling.su.se, mattias.heldner@ling.su.se

Abstract

This study of lexical stress in English is part of a series of studies, the goal of which is to describe the acoustics of lexical stress for a number of typologically different languages. When fully developed the methodology should be applicable to any language. The database of recordings so far includes Brazilian Portuguese, English (U.K.), Estonian, German, French, Italian and Swedish. The acoustic parameters examined are f_0 -level, f_0 -variation, Duration, and Spectral Emphasis. Values for these parameters, computed for all vowels, are the data upon which the analyses are based. All parameters are tested with respect to their correlation with stress level (primary, secondary, unstressed) and speaking style (wordlist reading, phrase reading, spontaneous speech). For the English data, the most robust results concerning stress level are found for Duration and Spectral Emphasis. f_0 -level is also significantly correlated but not quite to the same degree. The acoustic effect of phonological secondary stress was significantly different from primary stress only for Duration. In the statistical tests, speaker sex turned out as significant in most cases. Detailed examination showed, however, that the difference was mainly in the degree to which a given parameter was used, not how it was used to signal lexical stress contrasts.

Index Terms: speech prosody, lexical stress, English

1. Introduction

This paper presents one study in a planned series of studies describing and modelling the acoustics of word stress in a number of typologically different languages. The ultimate goal is to develop a model of analysis that may be applied to any language. At present, we have recorded data from Brazilian Portuguese, English, Estonian, French, German, Italian and Swedish. A first round of analyses has been published for Brazilian Portuguese [1, 2], Estonian [3], German [4] and Swedish [5, 6].

All languages that have contrastive word stress have primary stress. In some languages, the stress contrast is binary; a syllable can be stressed or unstressed. Many languages, like English and Swedish also have secondary stressed syllables, in which case three levels of stress must be considered.

Stress varies not only as a function of language but also with speaking style. Word list reading is likely to produce the most prototypical stress patterns, the ones we typically see described in lexica, whereas in spontaneous speech we are likely to find the acoustic stress correlates reduced or even missing. Phrase reading may be assumed to fall somewhere in between. We therefore also study the influence of speaking style on the acoustics of word stress. The speaking styles we

have investigated are the ones just mentioned, wordlist reading, phrase reading, and spontaneous speech.

The study of the acoustics of word stress goes back a long time. Classical studies in this area are those by Fry in the fifties e.g. [7]. In his study of English word stress, he found that f_0 -level and variation, vowel duration and vowel amplitude correlated with word stress but not all to the same degree. These early findings have largely been confirmed in a broad sense in studies of other languages like Polish [8], French [9], Swedish [10], and Spanish [11, 12]. We therefore have good reasons to consider these parameters as relevant in the acoustic description of stress. Amplitude (e.g. SPL) has however not turned out to correlate very well with stress level or stress perception. Another approach to what may also be interpreted as an “effort” measure, namely Spectral Balance (we prefer to call it Spectral Emphasis) has been shown to correlate with stress in studies of Dutch and American English [13-15]. Spectral Emphasis has also been shown to play a role in American English [16-18] and Swedish [19], but primarily in words pronounced with focal accent.

Based on the above studies and many more we have decided to approach the acoustics of word stress at this stage by analysing the following parameters: f_0 -level, f_0 -variation, Duration, and Spectral Emphasis.

2. Method

To minimise the influence of variation at the segmental level, we adopted a method that produced identical speech material in all speaking styles for a given speaker. This was obtained the following way. Each speaker was first recorded in a semi-spontaneous interview situation. They were free to choose the topic of the conversation. The interviews lasted 15–25 minutes. The recordings were manually transcribed using Praat TextGrids [20] and from these transcriptions we picked out 15–20 phrases where speech was fluent (i.e. no pauses, no false starts etc.) and which contained suitable target words of two or more syllables. Two manuscripts were prepared, one containing the target words in isolation, and one containing the corresponding phrases. Each word and phrase occurred three times in the lists, and the order between items was randomised. Two to four weeks after the interview session the speakers were recorded again, now reading the word and phrase lists based on their own spontaneous speech.

2.1. Speakers

The speakers (16 female; 15 male) were recruited among the students at University of Cambridge all speaking a variety of standard Southern English. They were all in the same age range (female speakers, 18–37 yrs., mean 22 yrs.; male speakers, 19–40 yrs., mean 26 yrs.).

2.2. Recordings

The recordings were made in a sound treated studio using *Sennheiser* HSP 4 cardioid headset microphones connected to a computer using the *Apple Logic Pro* recording software and an *M-AUDIO ProFire 2626* audio interface. Recordings were originally sampled at 48 kHz/16bit but they were down sampled to 16kHz/16bit for the acoustic analyses.

2.3. Parameters used in the acoustic analyses

In-data for fundamental frequency level and variation are median f_0 values within each vowel, expressed in semitones relative to 1 Hz. Median values were used to minimize the influence of measurement errors.

Fundamental frequency level is defined as the mean f_0 -level in semitones.

Fundamental frequency variation is defined as the Standard Deviation of f_0 -level in semitones.

Duration is measured in ms.

Spectral Emphasis is defined as the difference in dB between the Sound Pressure Level (SPL) of the full spectrum and the SPL of f_0 in each segment. For details see [21].

The use of the semitone scale for frequency means that we may expect the variation to be approximately the same for male and female speakers. The semitone scale also reduces skew. Using a log scale tends to make the distribution more normal. For this reason we express duration as $\text{Log}_2(\text{ms})$. Log-scales are thus used for all parameters.

2.4. Fixed factors used in the statistical analyses

Sex: Male, Female

Stress: Unstressed, Secondary, Primary

Style: Spontaneous, Phrase, Word

2.5. Extracting the parameter values

The parameter values were extracted using a *Praat* script specifically designed for the purpose. The script extracted a large number of parameters used in preliminary tests. Here we will only consider the parameters described in 2.3.

In preparation for applying the script, all recordings were manually transcribed in *Praat* TextGrids using four tiers; Phrase, Word, Segment and Stress level. The TextGrid files together with the sound files were used to extract the above-mentioned values segment by segment. The output from the script was a table where each line in the table contained the acoustic data segment by segment together with its phonological symbol, type (vowel/consonant), and stress level (primary, secondary, unstressed). Stress level annotation was based on a recognised pronunciation dictionary [22]. In the analyses presented here only the vowels in the target words have been considered.

2.6. Database used in the analyses

The procedure described in 2.5 gave us a database of parameter values for about 12000 vowels in total. The number of vowels per speaker group (male/female) is roughly the same (6000). The exact numbers vary slightly between parameters due to missing data, for example as f_0 could not always be reliably estimated.

3. Results

3.1. Fundamental frequency level

As we may see in Figure 1, the basic patterns of f_0 -level in the vowel as a function of stress level are the same for male and female speakers. For this particular parameter we already know, however, that male and female speakers will differ in overall f_0 -levels. The overall means for the females and males are 200 Hz and 110 Hz, respectively. This corresponds to a mean difference of 10.4 semitones between the groups. In order to make the analyses of between-subjects effects including *Sex* more meaningful, we equalized the mean f_0 -level by subtracting 10.4 semitones from each data point in the female data. A Univariate ANOVA using the equalized f_0 values as the dependent variable and *Stress*, *Sex* and *Style* as fixed factors showed significant main effects of *Stress* [$F(2,11643)=182.7$; $p < .001$], and *Style* [$F(2,11643)=35.0$; $p < .001$], as well as significant interactions between *Stress* and *Sex* [$F(2,11643)=25.4$; $p < .001$] and between *Stress* and *Style* [$F(4,11643)=17.0$; $p < .001$]. The main effect of *Sex*, as well as the interactions between *Sex* and *Style* and *Stress*, *Sex* and *Style* were not significant. The explained variance of this model is 7.3%.

Effects of stress level: When we examine the effects of stress level in more detail, we find that primary and secondary stressed vowels have significantly higher f_0 than unstressed vowels for both male and female speakers. For all data pooled, there is no significant difference between secondary and primary stressed vowels. If we analyse the speaker groups separately, there is no difference for female speakers. For male speakers the difference is significant. Furthermore, the significant interaction between *Stress* and *Sex* can be explained by the fact that the difference between unstressed and stressed (i.e. primary and secondary stressed pooled) is larger for the male speakers than for the female speakers; the differences are 2 semitones and 1 semitone, respectively.

Effects of speaking style: Looking at the effects of speaking style, we find that the effect is the same for male and female speaker, as the interaction between *Sex* and *Style* is not significant. However, the effect of stress is not the same in the different speaking styles, as shown by the significant interaction between *Stress* and *Style*. In particular, while primary and secondary stressed vowels have markedly higher F_0 than unstressed ones in both word lists and phrases, there is only a marginal effect of stress in the spontaneous data, although qualitatively the pattern is the same.

3.2. Fundamental frequency variation

Next, we turn to f_0 -variation as a function of stress level. A Univariate ANOVA with f_0 standard deviation (in semitones) as dependent variable, and the same independent variables as in the model for f_0 -level shows significant main effects of *Stress* [$F(2,11004)=6.8$; $p = .001$], *Sex* [$F(1,11004)=9.5$; $p = .002$] and *Style* [$F(2,11004)=5.2$; $p = .005$], as well as an interaction between *Stress* and *Sex* that only just reached significance [$F(2,11004)=3.1$; $p = .042$]. The other interactions are not significant. The explained variance for this model is 1%.

Figure 1: *Fundamental frequency level as a function of speaking style (word list, phrase reading and spontaneous speech) and stress level.*

Figure 2: *Fundamental frequency variation as a function of speaking style (word list, phrase reading and spontaneous speech) and stress level.*

Figure 3: *Vowel duration as a function of speaking style (word list, phrase reading and spontaneous speech) and stress level.*

Figure 4: *Spectral Emphasis as a function of speaking style (word list, phrase reading and spontaneous speech) and stress level.*

Effects of stress level: When we examine the effects of stress level in more detail, we find that the significant effect of *Stress* is primarily due to the greater f_0 -variation in unstressed vowels. Female speakers show significantly higher variation than male speakers (0.83 vs. 0.77 semitones). Looking at female and male speakers separately, we find that the effect is caused primarily by markedly more variation in the unstressed and secondary stressed vowels for the female speakers. This is also the explanation for the significant interaction between *Stress* and *Sex*.

Effects of speaking style: Examining the effects of speaking style, we find that the f_0 -variation is larger in wordlist reading compared to phrase reading or spontaneous speech (0.84 semitones vs. 0.75 semitones). Variation in spontaneous speech and phrase reading are virtually identical (0.75 semitones for both). The same pattern is observed if male and female speakers are analysed separately.

3.3. Duration

Next, we turn to duration as a function of stress level. A Univariate ANOVA with duration of the vowel (expressed as the binary logarithm of duration in ms) as dependent variable, and the same independent variables as in the other models shows significant main effects for *Stress* [$F(2,12511)=548$; $p < .001$], *Style* [$F(2,12511)=107$; $p < .001$] and *Sex* [$F(1,12511)=6.0$; $p = .015$]. In addition, there are significant interactions between *Stress* and *Style* [$F(2,12511) = 27.3$; $p < .001$] and *Stress* and *Sex* [$F(2,12511) = 3.7$; $p = .025$]. The explained variance for this model is 14.2%.

Effects of stress level: Stress level has a significant effect on vowel duration and for this parameter all three levels are significantly different. If we express the mean durations in milliseconds, they are 53, 66, and 79 ms for unstressed, secondary stressed and primary stressed vowels, respectively. The significant effect of *Sex* is due to longer mean durations for the female speakers. If we look at the speaker groups separately we may see that the difference is almost entirely caused by the markedly longer primary stressed vowels in the female group (85 ms vs. 78 ms) which explains the significant interaction between *Stress* and *Sex*.

Effects of speaking style: *Style* also makes a significant difference. The main difference is between word list reading (73 ms) and the other two styles, phrase reading (57 ms) and spontaneous speech (55 ms), but all comparisons are statistically significant. The interaction between *Stress* and *Style* is caused by the fact that the increase in duration as a function of stress level is almost identical for phrase reading and spontaneous speech, but markedly lower in word list reading.

3.4. Spectral Emphasis

Finally, we examine *Spectral Emphasis* as a function of stress level. A Univariate ANOVA with spectral emphasis in the vowel (in dB) as dependent variable, and otherwise the same independent variables as in the other models shows significant main effects of *Stress* [$F(2,12511)=416$; $p < .001$], *Style* [$F(2,12511)=158$; $p < .001$] and *Sex* [$F(1,12511)=416$; $p < .001$]. In addition, there is a significant interaction between *Sex* and *Style* [$F(2,12511) = 96.8$; $p < .001$]. The explained variance is 17.5%.

Effects of stress level: A Bonferroni Post-Hoc test shows significant differences between all three stress levels, although the difference between primary and secondary stressed is smaller than that between unstressed and stressed. Furthermore, if male and female speakers are tested separately, the secondary vs. primary difference is not significant for the female speakers. Male speakers have on average 1.2 dB higher average *Spectral Emphasis* than females. The increase in spectral emphasis from unstressed to primary stressed, however, is approximately the same (2.3 dB and 2.2 dB for male and female speakers, respectively).

Effects of speaking style: A Bonferroni test on the main effect of Style showed no significant difference between word list reading and phrase reading. Somewhat surprisingly, spontaneous speech had considerably higher spectral emphasis than the other two styles. If male and female speakers are analysed separately, the same pattern is found for both groups, but the difference is much greater for the male speakers than for the female speakers (3.4 dB vs. 0.5 dB). This is likely the explanation for the significant interaction between *Sex* and *Style*.

4. Discussion

4.1. Fundamental frequency level

Fundamental frequency level has been identified as an important acoustic correlate of stress in English and many other languages in studies from those by Fry [e.g. 7] and onwards. Our results agree with these claims although we arrive at the conclusion from a different angle.

The effect of stress on f_0 -level is the same for both speaker groups and in fact also in all speaking styles; there is an increase in f_0 from unstressed to primary stressed but the range is about 1 semitone smaller in the female data. We may say that the male and female speakers signal lexical stress the same way, but the female speakers use the f_0 cue to a somewhat lower degree. As we shall see below this pattern can also be found in other parameters.

4.2. Fundamental frequency variation

Fundamental frequency variation has also been proposed as an acoustic correlate of stress. However, our results show a minimal effect of stress level on f_0 -variation. We did however find differences between speaker groups, but only in the degree of variation. Furthermore, as the explained variance is only 1%, these results should be interpreted with caution.

4.3. Duration

Duration has also been proposed as a correlate of stress, and our results support these claims. Duration is the only parameter for which the values show a significant stepwise increase from unstressed to secondary stressed to primary stressed. These differences are consistent across speaker groups, as well as across speaking styles. Again, we find that the general patterns are the same for females and males although the extent to which duration is used varies if we look at details.

4.4. Spectral Emphasis

Spectral emphasis also turned out to be a reliable correlate of word stress, but primarily for making a distinction between

stressed and unstressed. Another observation is that the mean level for *Spectral Emphasis* is higher by 1.2 dB for the male speakers. The same difference has been observed in other languages [1–6]. The reason for this sex difference is not clear. One may ask if it is an artefact of the way it is calculated, but if so we cannot quite see how. It may of course also be the case that male speakers actually do produce higher *Spectral Emphasis*. This question needs further investigation.

Again, the patterns for male and female speaker are very similar, and the increase is the same for all speaking styles.

5. Conclusions

A significant observation is that the male and female speakers produce the stress contrast basically the same way. If we go down to details, we may observe some differences, primarily in the sense that male and female speakers use the same means but not always to the same degree.

This study shows English to be yet another language where Spectral Emphasis plays a significant role. And we may add that preliminary results from Italian point in the same direction. Most of the other studies of English have been on American English, and we cannot quite exclude the possibility that this variety of English is different from UK English in this respect.

Earlier studies of English have often ranked the acoustic parameters f_0 -level, Duration, Intensity. If we go by the degree of explained variance our ranking is quite different – Spectral Emphasis (17.5%), Duration (14.2%) and f_0 -level (7.3%).

Another observation is that the secondary stressed vowels are significantly different from primary stressed vowels only with respect to duration. For f_0 -level the situation is somewhat ambiguous as noted in 3.1. For f_0 -variation there is no support at all, neither globally nor for male and female speakers analysed separately. If our results are representative, they mean that the burden of signalling secondary stress contrastively depends almost entirely on duration. It should be mentioned, however, that in another study of word stress in English [18], the situation was the reverse. For primary and secondary stressed vowels, f_0 was found to be different. Duration, on the other hand, showed no difference.

6. 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.

We thank Francis Nolan, University of Cambridge for generously giving us full access to the recording studios when we collected the data. We also thank Chris Cummins and Toby Hudson for helping us to recruit speakers.

7. References

- [1] Barbosa, P. A., Eriksson, A., and Åkesson, J., “Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese,” in *Proc. Nordic Prosody XI*, 2013, pp. 97–106.
- [2] Barbosa, P. A., Eriksson, A., and Åkesson, J., “On the robustness of some acoustic parameters for signalling word stress across styles in Brazilian Portuguese,” in *Proc. Interspeech 2013*, 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*, 2014, pp. 232–235.
- [4] Behrens, J., "Die Prosodie des Wortakzentes in Abhängigkeit von Akzentlevel und Sprechstil," BA Thesis, Christian-Albrechts-Universität zu Kiel, 2013.
- [5] 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 XI*, 2013, pp. 127–136.
- [6] 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*, 2013, pp. 778–782.
- [7] 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.
- [8] Jassem, W. J., Morton, J., and Steffen-Bartóg, M., "The perception of stress in synthetic speech-like stimuli by Polish listeners," *Speech Analysis and Synthesis*, vol. 1, pp. 289–308, 1968.
- [9] Benguerel, A. P., "Physiological correlates of stress in French," *Phonetica*, vol. 27, pp. 21–35, 1973.
- [10] Fant, G. and Kruckenberg, A., "Notes on stress and word accent in Swedish," *STL-QPSR*, vol. 2-3, pp. 125–144, 1994.
- [11] Díaz-Campos, M., "The phonetic manifestation of secondary stress in Spanish," in *Hispanic Linguistics at the Turn of the Millennium: Papers from the 3rd Hispanic Linguistics Symposium*, Somerville, MA: Cascadilla, 2000, pp. 49–65.
- [12] Vargas-Calderon, R., "Analyse acoustique de l'accent de l'espagnol parlé au Costa Rica," in *Travaux de l'Institut de Phonétique de Strasbourg*, vol. 18, 1986, pp. 1–23.
- [13] Sluijter, A. M. C., "Phonetic Correlates of Stress and Accent," PhD thesis, Holland Academic Graphics, The Hague, 1995.
- [14] Sluijter, A. M. C., Shattuck-Hufnagel, S., Stevens, K. N., and van Heuven, V. J., "Supralaryngeal resonance and glottal pulse shape as correlate of stress and accent in English," in *Proc. ICPhS 95*, vol. 2, 1995, pp. 630–633.
- [15] 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.
- [16] Campbell, N. and Beckman, M. E., "Stress, prominence, and spectral tilt," in *Intonation: Theory, Models, and Applications*, Athens, Greece, 1997, pp. 67–70.
- [17] Campbell, N. and Beckman, M. E., "Accent, stress, and spectral tilt," *Journal of the Acoustical Society of America*, vol. 101, p. 3195, 1997.
- [18] Yuan, J., Isard, S., and Liberman, M., "Different roles of pitch and duration in distinguishing word stress in English," in *Proc. Interspeech 2008*, 2008, p. 885.
- [19] 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.
- [20] Boersma, P. and Weenink, D. (2014), "Praat: doing phonetics by computer" [Computer program]. Available: <http://www.praat.org/>
- [21] 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.
- [22] Wells, J. C., *Longman Pronunciation Dictionary*, 3rd ed. Harlow: Pearson, 2008.