The Acoustics of Word Stress in Swedish: A Function of Stress Level, Speaking Style and Word Accent

Anders Eriksson[1], Plinio A. Barbosa[2] and Joel Åkesson[1]

[1]Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden
[2]Speech Prosody Studies Group, Department of Linguistics, State University of Campinas, Brazil

`anders.eriksson@ling.gu.se, pabarbosa.unicampbr@gmail.com, joelakesson@gmail.com`

Abstract

The study presented here is one in a series of studies intended to describe the acoustics of word stress for several typologically different languages in a common framework. The idea is that, when fully developed the methodology should be applicable to any language in the same way regardless of prosodic type. The languages included in the present round of data collection and analyzes are Brazilian Portuguese, English, Estonian, French, Italian and Swedish. The acoustic variables examined here are F_0-level, F_0-variation, Duration, and Spectral Emphasis for all vowels in the data. All parameters are tested with respect to their correlation with stress level (primary, secondary, unstressed), speaking style (wordlist reading, phrase reading, spontaneous speech) and tonal word accent. The most robust results concerning stress level are found for Duration and F_0-variation. Speaking style turned out to play a minimal role. The only robust effect was found for duration which was longer in word list reading. Word accent had a significant effect on F_0-variation, and Duration.

Index Terms: speech prosody, word stress, tonal accent, Swedish

1. Introduction

This paper presents one study in a planned series of studies aiming at describing and modelling the acoustics of word stress in a number of typologically different languages. The ultimate goal is to develop a model for this type of analysis that may be applied to any language. At the present stage we have recorded data from English, Brazilian Portuguese, Italian and Swedish. Data collection is in progress for two additional languages; French and Estonian. A first set of analyses has been completed for Brazilian Portuguese [1] and Swedish. In the present databases we have recordings from 10 speakers per language (5 female and 5 male). In order to improve precision in our models and predictions we are, however, in the process of collecting more data. This work is planned to take part during the latter part of this year.

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. Text reading may be assumed to fall somewhere in between in this respect, but this has not been very well studied. In the present series of studies we also study the acoustics of word stress as a function of speaking style. The speaking styles we have investigated are the ones just mentioned, wordlist reading, phrase reading, and spontaneous speech.

Stress level is another parameter to consider. All languages that have 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. Where this is the case, three levels of stress must be considered. In Swedish all compounds have secondary stress and many singleton words as well. Some languages may have other contrastive prosodic means that may compete with stress perceptually and may influence the acoustic realisation of word stress. Swedish is a language where this is the case. Swedish is not a tone language, in the strictest sense of that concept, but it has a tonal accent that is contrastive in a limited vocabulary of singleton words and always present in compounds. For Swedish it is therefore necessary to consider tonal word accent (where applicable) when analysing the acoustics of word stress. In the following the two contrasting accents will be referred to as *Accent 1* and *Accent 2*. The latter is the accent type used in compounds.

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. [2]. 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 by and large been confirmed in a broad sense in studies of other languages like Polish [3], French [4], Swedish [5], and Spanish [6], [7]. 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 [8], [9].

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, Vowel Duration, and Spectral Emphasis.

2. Method

With so relatively few speakers per language we need to take measures to minimise variation as far as possible. One way to do so is to have 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. An interview lasted 15–25 minutes. These recordings were then transcribed using Praat TextGrids [10] and from these transcriptions we picked out 15–20 phrases were speech was fluent (i.e. no pauses, no false starts etc.). Another requirement when choosing the phrases was that they should contain suitable target words of three or more syllables. Compounds were accepted in order to study the influence of the above mentioned tonal accent. Based on the transcriptions, 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 were recruited among the students at University of Gothenburg all speaking a variety of standard Gothenburg Swedish (a requirement). They were all in the same age range (female speakers, 26–33 yrs, mean 30 yrs; male speakers, 26–37 yrs, mean 30 yrs). They were all, with one exception (a PhD), university students at the time of recording.

2.2. Recordings

The recordings were made in a sound treated recording studio using *Sennheiser* HSP4 EW-3 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/16 bit but for the acoustic analyses they were later down sampled to 16kHz/16bit.

2.3. Parameters used in the acoustic analyses

Fundamental frequency level is here defined as the F_0 median in order to minimize the influence of outliers. The median is measured in Hz.

Fundamental frequency variation is defined as the Standard Deviation of F_0 in semitones. The use of semitones is motivated by the fact that using this measure we may expect the variation to be approximately the same for male and female speakers.

Segment duration is measured in ms.

In these analyses we used a simplified version of the *Spectral Emphasis*.

$$\text{Spectral Emphasis (dB)} = \text{SPL}_{full} - \text{SPL}_0$$

SPL_{full} is the SPL of the full spectrum in the segment and SPL_0 is the SPL of F_0 in that segment. For further details see [11]

2.4. Extracting the parameter values

The parameter values were extracted using a Praat script, LexStressProcessing, 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 the previous paragraph. In preparation for the application of the script, all recordings were further transcribed. In this phase, the transcriptions at the word and phrase level used for the manuscripts were modified by adding two new tiers, a segment tier using a phonological representation of the speech sounds and a stress tier where the degree of stress – primary, secondary and unstressed – was marked. For Swedish the stress tier coding included information about whether the sound occurred in an Accent 1 word or an Accent 2 word. The TextGrid files together with the sound files were now used by the LexStressProcessing script to extract the above mentioned values segment by segment. The output from the script was a table were 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). In the analyses presented here only the vowels in the target words have been considered.

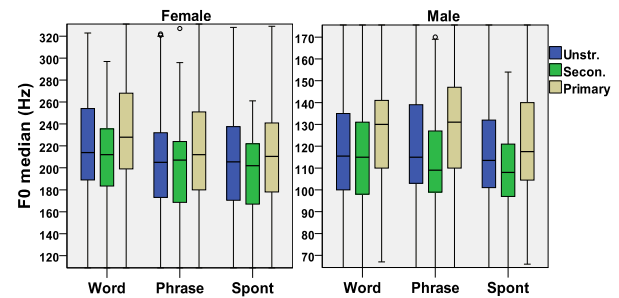# 3. Results

3.1. Fundamental frequency level

Figure 1: *Fundamental frequency median as a function of speaking style and stress level.*

As we may see in Figure 1, the basic patterns are the same for male and female speakers.

Effects of stress level: Primary stressed vowels have higher F_0 than secondary stressed and unstressed vowels for both male and female speakers. The differences are statistically significant (Univariate Anova, Post hoc LSD; $p < 0.05$) except for primary stressed vowels compared to unstressed vowels in spontaneous speech where the differences are only trends (p=0.30 and 0.26 for female and male speakers respectively). Unstressed vowels have higher F_0 than secondary stressed vowels for both male and female speakers and all speaking styles, but the differences reach statistical significance only for male speakers in the phrase reading condition.

Effects of speaking style: For female speakers vowels have significantly higher F_0 ($p < 0.05$) in word list reading than in phrase reading or spontaneous speech. Phrase reading and spontaneous speech do not differ significantly. For male speakers F_0 is higher in word lists and phrase reading than in spontaneous speech. No other differences are significant.

Summary: Primary stressed vowels have higher F_0 than secondary stressed or unstressed vowels for both speaker groups and all speaking styles although not all differences reach statistical significance. F_0 level is highest in the word list condition, while no clear-cut differences appear between the other two speaking styles.

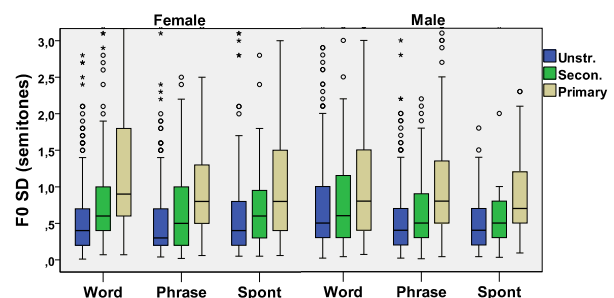## 3.2. Fundamental frequency variation

Figure 2: *Fundamental frequency variation as a function of speaking style and stress level.*

Effects of stress level: There is a consistent effect of stress level for all speaking styles and both speaker groups. It also follows an expected pattern – Primary > Secondary >

Unstressed. Many, but not all, of the differences are statistically significant at the 0.05 level, however. Primary stressed vowels show significantly more variation than unstressed vowels in all cases except for spontaneous speech produced by male speakers. The differences between secondary stressed and unstressed vowels are significant for word list reading but only trends in the other cases. Not surprisingly, the trends are weakest for the spontaneous speech data. A partial explanation for the lack of a clear significance in all cases is the relatively small number of speakers in each group (5 speakers). If data for both groups are pooled the Primary > Secondary > Unstressed hierarchy is significant except for secondary vs. unstressed in spontaneous speech.

Effects of speaking style: There is no significant effect at all of speaking style for any of the stress types (primary, secondary, unstressed) in either speaker group (male/female).

Summary: For this parameter we may observe a very consistent variation as a function of stress level. The variation is statistically significant in all cases but one if data from male and female speakers are pooled. For speaking style, in contrast, no such relationship may be observed.
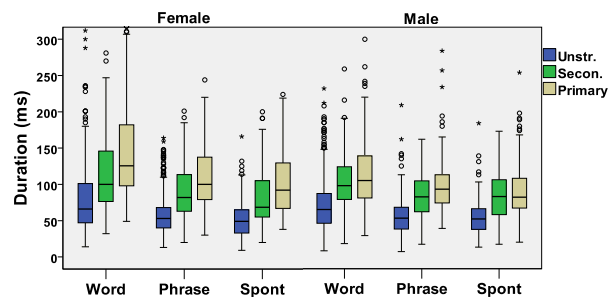
3.3. Segment duration

Figure 3: *Vowel duration as a function of speaking style and stress level.*

Effects of stress level: The Primary > Secondary > Unstressed hierarchy observed for F_0-variation is even more pronounced here. All differences as a function of stress level are highly significant (GLM, Univariate, $p < .001$) in all speaking styles except for Primary vs. Secondary stress for spontaneous speech and male speakers where it is only a trend ($p = .060$).

Effects of speaking style: There is also a consistent effect of speaking style. There are no significant differences between phrase reading and spontaneous speech in any of the comparisons. Word list reading, however, produces longer durations than phrase reading or spontaneous speech for all stress levels and both speaker groups ($p < .001$).

3.4. Spectral Emphasis

A first observation (Figure 4) is that Spectral Emphasis is higher for male speakers. The difference is highly significant (Anova, $p < .001$). There is also less variation in the male data.

Effects of stress level: The patterns are very similar for male and female speakers. Spectral Emphasis is also very similar for primary stressed and secondary stressed vowels whereas it is lower in unstressed vowels than in the other two types. For the male speakers this impression is confirmed by the statistical analysis. There is no significant difference between primary stressed and secondary stressed vowels for any of the

speaking styles. Unstressed vowels in contrast, have significantly lower Spectral Emphasis than primary stressed and secondary stressed ones ($p < .005$) in all cases. There are trends in the same direction for the female data, but due to the considerably larger variation few differences reach statistical significance. We see no strong reasons to believe that this is a reflection of sex differences. The considerably lower over all levels for female speakers on the other hand may well be a sex difference, but this needs to be further investigated.

Effects of speaking style: For Spectral Emphasis the influence of speaking style produces a confusing picture. As may be observed in Figure 4, Spectral Emphasis seems a bit higher in spontaneous speech for female speakers but somewhat lower for male speakers. The observation receives some statistical support for female speakers ($p < .05$ in most comparisons). For male speakers the support is limited to trends in some cases.
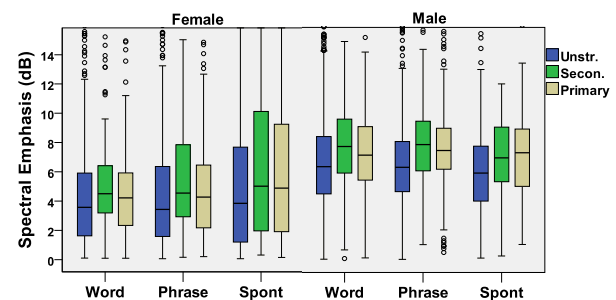

Figure 4: *Spectral Emphasis as a function of speaking style and stress level.*

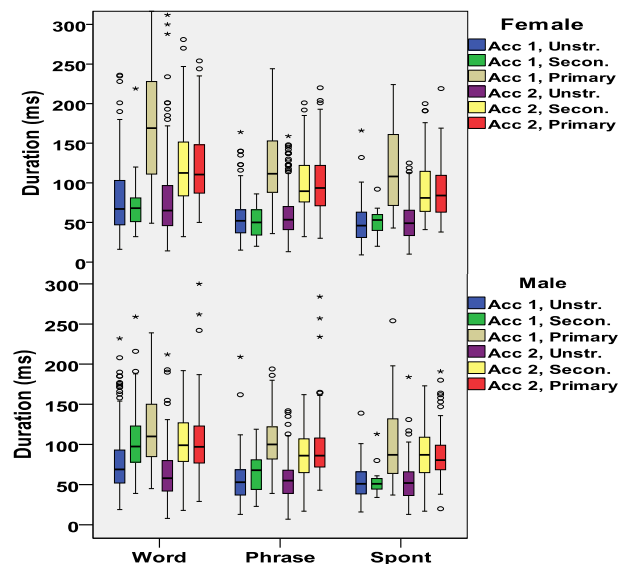## 4.   The influence of tonal accent

Figure 5: *Vowel duration as a function of speaking style, stress level and tonal accent.*

Due to space limitations we will only show a diagram for the parameter where the effect of tonal accent is most pronounced and refer to this picture as a general reference when discussing the other parameters.

If we look at the data for female speakers in the Figure 5, we may observe a rather striking difference between the two

accent types. For Accent 1 words, unstressed and secondary stressed vowels do not seem to differ in duration, whereas primary stressed ones are distinctly longer. The observed pattern is statistically highly significant. Unstressed and secondary stressed vowels are not significantly different, whereas primary stressed vowels are significantly longer than both the other types in all cases ($p < .001$). For Accent 2 words the pattern is reversed. Here the primary and secondary stressed vowels have similar durations while the unstressed ones are considerably shorter. The statistical support is equally solid in this case. For male speakers the patterns are the same but significance levels are somewhat lower.

The graphical picture for F_0-variation is strikingly similar to the one presented in Figure 5. Due to more variation in the data, however, the statistical analysis does not lend equally strong support in this case. In 6 of the 12 cases the statistical support is at the same level as for duration and in another two there are trends in the expected direction. If only Accent 1 is considered then the statistical support for the pattern shown for duration is at the same level for F_0-variation. It is the greater variation for Accent 2 data that make these results less robust.

We may summarise the findings so far the following way. Primary stressed vowels are longer/more varied than unstressed ones in both accent condition, whereas the status of secondary stressed vowels vary with accent type. In Accent 1 words secondary stressed vowels do not differ significantly from unstressed ones. In Accent 2 words it is primary and secondary stressed vowels that do not differ.

For the other parameters analysed here, F_0-level and Spectral Emphasis, we found little that systematically distinguishes Accent 1 from Accent 2. For F_0-level, primary stressed vowels have significantly higher F_0 ($p < .05$) in Accent 2 words for both groups, but other than that no systematic pattern may be identified. For Spectral Emphasis some cases show a pattern similar to that for duration and F_0-variation, but there is no statistical support for any definitive conclusions. We would like to leave the question open, however, that there may also be some systematic influence of accent type on Spectral Emphasis but that in this investigation it is obscured by too much noise in the data. Future studies may clarify this point.

5. Discussion

If we first consider the three global factors examined in the present study – stress level, word accent and speaking style – we have seen that there is a strong influence of stress level on all parameters. For F_0-level it is only the stressed vowels that stand out as different, but for those the pattern is consistent and significant. The absolute differences are small, however, and one may question if they could play any significant role in the perception of stress. Word accent shows a very marked effect on secondary stressed vowels for duration and F_0-variation but only a very marginal effect on F_0-level and a somewhat confusing picture for Spectral Emphasis. The influence of speaking style is minimal, not quite consistent and seldom statistically significant.

Duration is the parameter that shows the most stable pattern and also the highest degree of statistical significance. This comes as no surprise. Swedish is a quantity language where quantity, only present in stressed syllables, is signalled primarily by duration. In such a language it is important for speakers to maintain consistency in the production of quantity.

We also know from studies of other types of highly automated motor patterns that once established they tend too be very robust ([12], [13]).

For F_0-variation the pattern is the same as for duration, the only difference being that there is a little more variation in the data and therefore the statistical support is not quite at the same level.

The results for Spectral Emphasis are not quite as straightforward as for duration and F_0-variation. A first observation is that Spectral Emphasis is significantly higher for the male speakers. This may be due to a real sex difference in this respect but we have no independent evidence that this is the case and it has not, as far as we are aware, been reported in comparable investigations. This is thus a discovery that needs further investigation. Within the speaker groups, the patterns are very similar. The general tendency is that Spectral Emphasis is the same for primary stressed and secondary stressed vowels and that both are higher than the unstressed ones. The tendency is statistically significant for the male speakers but only trends in most cases for the female speakers. We have no explanation for the seemingly contradictory effect of speaking style for male and female speakers.

6. Conclusions

Our assumption that the parameters we suggest for a detailed look at the influence of stress level, speaking style and tonal accent are very useful in this respect are confirmed by the present results. They do not all, however, carry the same weight for any given language. For Swedish it seems that duration is the factor where the effect of stress level and tonal accent is most consistent. As explained above, this is hardly surprising since Swedish is a quantity language. The fact that F_0-variation shows an almost equally stable pattern is, however, not possible to predict from any phonological considerations and it has not, as far as the present authors are aware, been shown in any previous investigation of Swedish word stress.

In most comparisons there was little or no significant effect of speaking style. This is particularly noteworthy when phrase reading and spontaneous speech are compared. It is often claimed that phrase reading is not representative of spontaneous speech but this idea receives no support in the present study. The only statistically robust effect of speaking style was for Duration where word list reading produced significantly longer durations.

We are aware that more precise prediction may be possible given more data and that interesting details may be obscured by noise in the data. A larger database is likely to clarify this point and we are at present collecting more data for the languages included in the research programme. At the time of writing, we have recorded 20 additional speakers for Swedish but the data have not yet been analyzed. Additional recordings of the other languages are in the pipeline. We expect this work to be completed towards the end of 2013.

8. References

[1] Barbosa, P.A., Eriksson, A. and Åkesson, J., "On the robustness of some acoustic parameters for signalling word stress across styles in Brazilian Portuguese", These proceedings, 2013.

[2] Fry, D.B., "Duration and intensity as physical correlates of linguistic stress", *J. Acoust. Soc. Am.,* 27(4), 765–768, 1955.

[3] 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,* 1, 289–308, 1968.

[4] Benguerel, A.P., "Physiological correlates of stress in French", *Phonetica***,** 27**,** 21–35, 1973.

[5] Fant, G. and Kruckenberg, A., "Notes on stress and word accent in Swedish", *STL/QPSR,* (Vol. 2–3) 125–144, 1994.

[6] Vargas-Calderon, R., "Analyse acoustique de l'accent de l'espagnol parlé au Costa Rica", Travaux de l'Institut de Phonétique de Strasbourg, 18, 1–23, 1986.

[7] Díaz-Campos, M. "The Phonetic Manifestation of Secondary Stress in Spanish", In Campos, *et al*., *Hispanic Linguistics at the Turn of the Millennium: Papers from the 3rd Hispanic Linguistics Symposium*, pp. 49–65. Somerville, MA.: Cascadilla, 2000.

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

[9] Sluijter, A.M.C. and V. J. van Heuven, V.J., "Spectral balance as an acoustic correlate of linguistic stress", *J. Acoust. Soc. Am.,* 100(4 Pt. 1), 2471–2485, 1996.

[10] "Praat: doing phonetics by computer" (Version 5.1.37) [Computer program], Online: http://www.praat.org.

[11] Traunmüller, H. and Eriksson, A., "Acoustic effects of variation in vocal effort by men, women, and children", *J. Acoust. Soc. Am*., 107(6), 3438–3451, 2000.

[12] Viviani, P., and Terzuolo, C.A., "The organisation of movement in handwriting and typing", In B. Butterworth (Ed.), Language Production in Non-Speech Modalities (pp. 104–146). London, Academic Press, 1983.

[13] Eriksson, A., and Wretling, P., "How flexible is the human voice? - A case study of mimicry", In G. Kokkinakis, N. Fakotakis, E. Dermatas, (Eds), Eurospeech '97 Proceedings, Vol. 2, pp. 1043–1046). Rhodes, Greece, 1997.