Perception of Sentence Stress in English Infant Directed Speech

Sofoklis Kakouros [1], *Okko Räsänen* [1]

[1] Department of Signal Processing and Acoustics, Aalto University, Finland
sofoklis.kakouros@aalto.fi, okko.rasanen@aalto.fi

Abstract

Various studies have examined the acoustic features in infant directed speech (IDS) and adult directed speech (ADS). However, there are few speech corpora with prominence annotation from multiple listeners or analysis of the acoustic properties of the stressed versus unstressed words, most studies and corpora focusing on syllabic stress. In order to fill this gap, the current study analyzes the acoustic properties of sentence stress in a corpus of English IDS. More specifically, the work is one of the first analyzing IDS as perceived by adult listeners, providing inter-annotator agreement ratings and an analysis of the acoustic correlates of sentence stress with regard to the most important prosodic features encountered in the literature: fundamental frequency, intensity, word duration, and spectral tilt. The analysis shows that all of the analyzed features correlate with the perception of stress, indicating that the sentential prominence in IDS is conveyed by similar acoustic characteristics that are known to be relevant for stress perception in ADS.

Index Terms: sentence stress, prosody, infant directed speech, inter-rater agreement, annotation

1. Introduction

Stress is a prosodic phenomenon that can be generally defined as an accentuation of syllables within words, or of words within sentences [1]. In overall, prosodic phenomena typically manifest as contrastive changes in the acoustic features of speech that take place over different domains such as word, phrasal, and sentential level [2]. In this paper, we focus on the acoustic features of sentence stress. Many studies have earlier examined the acoustic properties of lexical and sentential stress in various languages in adult-directed speech (ADS) (see, e.g., [3,4,5]) but few in infant-directed speech (IDS) (see, e.g., [6]). Moreover, few speech corpora with prominence annotations are currently available (see, e.g., [7]). In this work we collected prominence data and examined stress in English IDS as reflected in the perception of adult listeners by analyzing the most important acoustic correlates found in the literature. We also analyzed inter-rater agreement levels in the stress perception task. The results show that sentence stress in IDS is conveyed in a similar manner as in ADS but with seemingly exaggerated changes in the features.

1.1. Acoustic correlates of sentence stress

Sentence stress is a universal property of speech where one or more words within the boundaries of a sentence receive a special emphasis. This typically implies that the speaker applies a modulation on the prosodic coding which can be independent of the linguistic content of the communicated message. On the listener's side, this translates into the perception of saliency in the message which may serve a number of different communication purposes from the speaker. For instance, it could convey the part of the sentence which carries the most important information that the speaker wants to communicate, or it could convey linguistically motivated information. It has been found, for example, that words receiving sentence stress have typically faster reaction times from the listeners suggesting that they are processed faster (see, e.g., [8]). Therefore, sentence stress, and stress in general has an impact on the perceptual processing of speech.

The perceptual effects of stress are present in both IDS and ADS, with however, some differences. In IDS for example, the increased perceptual clarity of the phonetic units in speech may assist the infants in language learning (see, e.g., [9], and references therein). Nonetheless, there are perceptual effects which are common in both IDS and ADS, such as attracting and holding the focus of attention [9]. What is more, IDS is considered to have some similarities with ADS clear speech [9]. Although there are certain similarities and differences in the perceptual processing of IDS and ADS, the acoustic features which are driving these changes are the same in both.

Specifically, a number of studies have examined the acoustic correlates of stress in different languages (see, e.g., [10-19]) giving consistent indications of their connection to stress perception. It is important to note at this point that, although the specific realization of stress might differ from one language to another, the acoustic correlates of stress seem to be universal (see, e.g., [20]). Therefore, the same set or subset of acoustic correlates may be used across languages in order to signify stress. In this regard, the following prosodic correlates are commonly used: (i) fundamental frequency, (ii) intensity, (iii) duration, and (iv) spectral tilt.

For instance, studies such as that of Fry [10] and Lieberman [12] showed that higher fundamental frequency (F0) in syllables is among the most relevant indicators of stress. In another study, Sluijter et al. [15] investigated several acoustic features in Dutch and their reliability as cues to stress. They found that duration and spectral tilt were the strongest indicators of stress. Finally, with regard to the intensity and duration, there is strong evidence of their importance as a cue to stress from the literature, with duration having the strongest consensus among the studies (see, e.g., [16], and references therein). Intensity seems also to play an important role in the perception of stress and is supported by studies such as that of Kochanski et al. [19].

In this paper, we will examine the effect of all four acoustic correlates on the perception of sentence stress in IDS. We will first describe the used material and methods. These are followed by the results, a short discussion, and the most important conclusions from this work.

2. Material and methods

The material used in this work consists of speech stimuli taken from a corpus on infant directed speech and an extensive data collection scheme for annotating sentence level stress.

2.1. Data collection

Data were collected in order to create a reference annotation of sentence level stress against which the acoustic parameters in the speech material would be compared. For this purpose, IDS material was used in a listening experiment designed to collect sentence stress annotations through an interactive tool.

2.1.1. Speech material

The speech stimuli in this study were taken from the CAREGIVER Y2 UK corpus [21]. The style of speech in CAREGIVER is acted IDS spoken in continuous UK English, corresponding to a situation where a caregiver is talking to a child regarding a number of jointly-attended objects and events in a shared interaction scene, but recorded in high-quality within a noise-free anechoic room. All talkers were either parents or had other experience with young infants, therefore this material was assumed as equivalent to typical IDS. In addition to a set of 50 unique keywords, there are a number of verbs and function words used in the surrounding carrier sentences of the corpus, yielding a total vocabulary of 80 words. The talkers were not separately instructed on the use of prosody or stress beyond that they were asked to read the text prompts, paired with visual stimuli, as they would talk to their own child (see [21], for details).

In overall, the "main talker section" of CAREGIVER contains 2397 sentences. A subset of 300 unique utterances was chosen for the listening tests from one male and one female talker (*Speakers 3* and *4*), yielding a total of 600 sentences. All single-word sentences were excluded from the data and there were 5.9 words per sentence on average.

2.1.2. Participants

A total of thirteen subjects (6 female, 7 male) participated in the listening experiment. The participants were recruited among the students and personnel of Aalto University and University of Helsinki, Finland. In the current study, we limit the analysis to the group of nine native (L1) Finnish speakers (6 female, 3 male) as the remaining four speakers had different L1 background. In this group, ages ranged between 26-30 years and English represented the L2 for all listeners. Each listener in the experiment reported to be a professional-level English speaker.

2.1.3. Procedure

A tool with a graphical user interface (GUI) was created in Matlab for the annotation. The function of the GUI was to play each utterance through headphones, display the list of spoken words in a temporally ordered manner, and prompt the user to select the words that were perceived as stressed using a mouse as a controller. In each utterance, the user could select zero or more words as stressed with the maximum being the total number of words spoken. Finally, the users could listen each sentence as many times as they wished.

The listening tests were carried out in a sound-isolated listening booth using Sennheiser HD650 headphones fed through Motu Ultralink MK3 audio interface. The listeners were able to take a break any time during the annotation process but were requested to take at least one following the completion of the first part (*Speaker 3*) of the task. On average, the annotation procedure lasted approximately 1.5 hours per listener.

2.2. Feature selection and extraction

Four features were used in the data analysis, namely: (i) signal energy, (ii) F0, (iii) spectral tilt, and (iv) word duration. For energy, F0, and spectral tilt the speech data were first downsampled from 44.1 kHz to 8 kHz. F0 contours for voiced segments were extracted for each utterance using the YAAPT-algorithm [22] with 25 ms window length and 10 ms frame shift. In order to ensure temporal continuity of the signal, the F0 contours during the unvoiced sections were generated by linear interpolation of the neighboring voiced F0 values [23]. Signal energy was calculated using the same window size and frame shift based on Eq. (1):

$$E = \log_{10}\left(\sum_{n=n_1}^{n_2} |x[n]|^2\right) \quad (1)$$

Spectral tilt was computed using the same windowing parameters and by taking the first Mel-frequency cepstral coefficient (MFCC) of each window (see, e.g., [24]). Finally, word duration was taken directly from the CAREGIVER Y2 UK corpus where the temporal boundaries of each word in the utterances, t_1 (word start) and t_2 (word end), were extracted from the word-level transcriptions.

In order to ensure comparability across talkers and utterances, all features were min-max normalized per utterance according to

$$f_\psi'(t) = \frac{f_\psi(t) - \min(f_\psi)}{\max(f_\psi) - \min(f_\psi)} \quad (2)$$

where $f_\psi(t)$ represents the value of feature ψ at time t (see [25]) and the min and max are computed across the entire utterance.

The main parameters which were then used in the analysis of the data were calculated over the duration of each individual word (see [4]). Based on the literature, the most common acoustic variables are the mean, median, and variation of the features (see, e.g., [3,4,5,26]). Furthermore, the study of maximum and feature change might also give meaningful information (see, e.g., [18]). For consistency, we calculated the same measures over all features in order to gain an understanding of their behavior. Therefore, we included the following in the analysis: (i) feature change computed according to Eq. (3), (ii) maximum feature value during the word, (iii) median or mean during the word, and (iv) variation calculated as the standard deviation of the feature during the word.

$$f_\psi^{ch} = \max\{f_\psi'(t)\} - \min\{f_\psi'(t)\}, t \in [t_1, t_2] \quad (3)$$

2.3. Inter-annotator agreement measures

In order to measure inter-annotator agreement rate in the listening test and compare our results with other similar studies in stress or prominence detection, the standard Fleiss kappa [27] measure was used. In essence, the Fleiss kappa measures the degree of agreement between two or more annotators on a nominal scale of $\kappa \in [-1,1]$. Fleiss kappa is a generalization of the Cohen's kappa statistic [28] which is applicable only on pairs of annotators and therefore could not be used in the context of this study.

The Fleiss kappa yields $\kappa = 0$ if the number of agreements is equal to what is expected based on chance-level co-occurrences in the data and $\kappa = 1$ if all annotators fully agree on all annotated tokens. In this work, Fleiss kappa was measured on the word-level. In particular, for each word occurring in the test set, a binary decision between non-stressed and stressed was considered. The overall agreement rate on the words in the test set was then used as the primary measure in the analysis. The overall kappa was measured then across all nine annotators, but also in a pair-wise manner for each possible pair of annotators in order to understand differences between listeners in the task.

Finally, in the analysis of the individual acoustic features with regard to the perception of sentence stress from the annotators, a single reference set was generated from the nine annotations. The set contained the majority agreement of the annotators where for each individual word in the test set a binary decision was made on whether the word was stressed or unstressed. Therefore, all words receiving five or more stress votes were marked as stressed and the rest as unstressed.

Table 1: *Fleiss kappa inter-annotator agreement rates.*

	Male speaker	*Female speaker*	*Both speakers*
Male listeners	0.39	0.49	0.44
Female listeners	0.30	0.46	0.38
All listeners	0.32	0.47	**0.40**

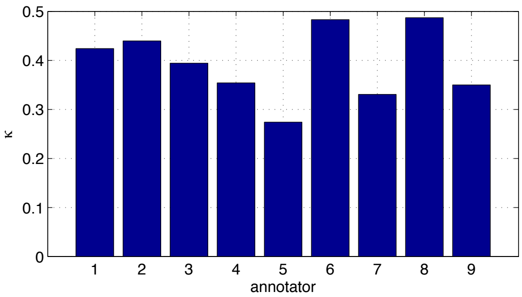

Figure 1: *Mean pair-wise agreement rates between annotators (mean agreement of each annotator with respect to the other 8 annotators).*

3. Results

3.1. Annotation analysis

Data from nine subjects over a total of 600 stress-annotated utterances were used in this analysis. In overall, there were 3569 words in the sentences where 26.3% (N = 938) of them were perceived as stressed by the majority of the annotators and 73.7% (N = 2631) as unstressed. A total of 70 unique words occurred in the utterances and 59 of them were marked as stressed at least once. The most frequent temporal locations for stressed words, when split in four quartiles over the length of a sentence, were the second and last quartiles (see Fig. 2).

The overall kappa across all annotators was κ = 0.4 (see also Table 1 for more details) which is exactly on the boundary between *fair* and *moderate* agreement level as derived by the Landis and Koch scale [29]. The kappa value translates into a mean word-level agreement rate of 84.9% on stress/non-stress decisions. The overall agreement rates observed are also basically equivalent to those in two other studies on prominence perception with American English [30,31].

Regarding the pair-wise agreement rates between the annotators, the average level was κ = 0.39 with a standard deviation of σ = 0.07 where the minimum mean pair-wise rate was 0.27 and the maximum 0.49 (see also Figure 1), indicating a certain degree of variation between the annotators. As for the statistical reliability, there was a significant difference between the agreement rate on male and female talkers by female listeners and for the pool of both male and female listeners ($p < 0.01$, unpaired t-test).

3.2. Fundamental frequency

Before comparing the measured features, the vocabularies used for the computation were matched between the stressed and unstressed words. This is because the function words (“*do*”, “*a*”, “*an*”, ”*the*”, “*at*”, “*he*”, “*she*”, “*you*”, “*I*”, “*and*”, “*is*”, “*has*”, “*have*”) represented 13 (18.6%) out of the 70 unique words in the test set but as many as 1521 (42.6%) out of 3569 of all word tokens. Moreover, only 6 unique function words were marked as stressed for a total of 38 times. Therefore, we used in our analysis all stressed words and two sets of unstressed words: one containing all unstressed words and another with only the unstressed content words.

The first feature analyzed in the collected data was F0. The four measures examined can be seen in Figure 3 where each histogram contour is normalized by the sum of frequency counts. Based on the results, F0 change and max seem to be the most descriptive in distinguishing between stressed and unstressed words ($p < 0.01$, unpaired t-test between stressed and both sets of unstressed words), since, as can be observed, many of the stressed words had a peak near one (N = 235, for F0 change, and N = 521, for F0 max). It seems that during a word, greater F0 change and F0 maxima result into the perception of prominence. It is important to note here that since the F0 is normalized across each utterance, 1 reflects the maximum F0 during that utterance and 0 the minimum.

3.3. Energy

Next, the effect of energy was examined and a consistent connection between the magnitude of energy and perception of stress was found. Specifically, stressed words in our data seem to be accompanied by high energy in the signal whereas the unstressed have an overall lower energy. As can be seen from Figure 4, this is reflected in most measures used in the study where, for instance, stressed words had higher mean energy than the unstressed ones (peak at 0.67 with N = 116 for the unstressed without function words and at 0.74 with N = 124 for stressed, $p < 0.01$, unpaired t-test).

3.4. Duration

When comparing word durations, the selection of vocabularies is particularly important. As the function words in our data are shorter in duration (μ = 0.1 s, σ = 0.06 s) when compared to the content words (μ = 0.38 s, σ = 0.14 s), they would bias the duration distribution. Therefore, only the histograms of the durations of the content words are plotted in Figure 5.

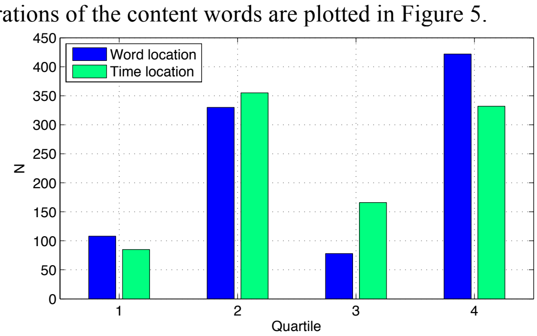

Figure 2: *Location of stressed words in the utterances (divided into quartiles). The blue bars show the ordinal position of the word in the sentence while the green bars show the temporal position with respect to utterance duration.*

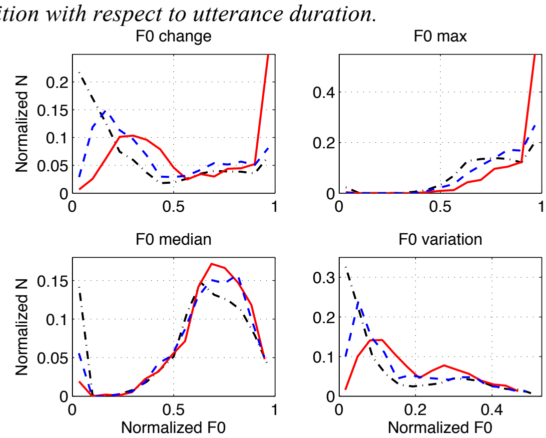

Figure 3: *Histograms with normalized frequency counts for F0. Red solid line: all stressed words, black dash-dotted line: all unstressed words, blue dashed line: unstressed without function words.*

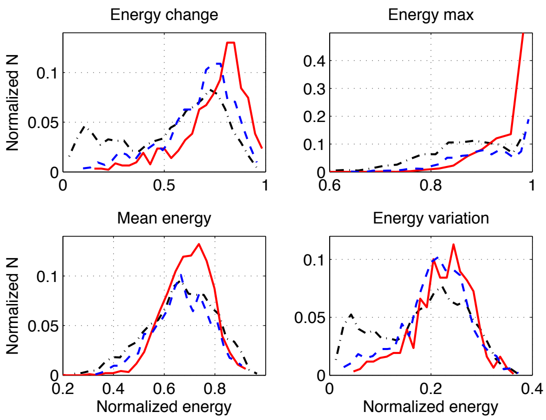

Figure 4: *Histograms with normalized frequency counts for energy. Red solid line: all stressed words, black dash-dotted line: all unstressed words, blue dashed line: unstressed without function words.*

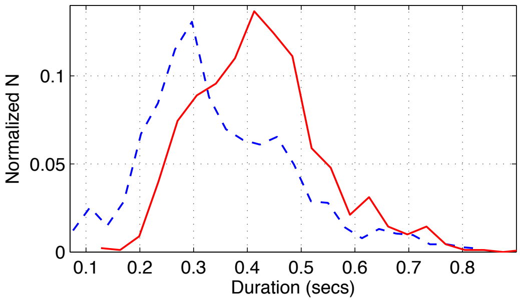

Figure 5: *Histograms with normalized frequency counts of the content words for duration. The red solid line represents the words perceived as stressed while the blue dashed line represents the unstressed.*

The two histograms clearly show that the stressed words are typically longer in duration, having an average duration of $\mu = 0.42$ s and standard deviation of $\sigma = 0.12$ s, while the unstressed words have $\mu = 0.35$ s and $\sigma = 0.14$ s, respectively ($p < 0.01$, unpaired t-test). As a reference, the corresponding distribution including also the function words has the same mean and standard deviation for the stressed words but notably lower $\mu = 0.21$ s and $\sigma = 0.16$ s for the unstressed words.

3.5. Spectral tilt

Finally, spectral tilt was examined and the results can be seen in Figure 6. It is important to note here that higher values of the normalized tilt (values closer to 1) translate into a steeper negative slope in the spectrum. Similarly with the energy feature, most measures computed for the spectral tilt suggest a connection with the perception of stress. For instance, stressed words appear to have higher tilt change (peak at $N = 188$) as compared to the unstressed (peak at $N = 299$ for all unstressed and $N = 172$ for the unstressed without function words) where an unpaired t-test shows a significant difference ($p < 0.01$). Therefore, greater changes in the steepness of the spectral slope during a word seem to translate into an increased perception of prominence.

4. Discussion and Conclusions

In this work we studied how the acoustic correlates of stress in IDS manifest in the perception of sentence level stress by adults. We examined four features and our results showed that all of them had a consistent effect in discriminating between stressed and unstressed words.

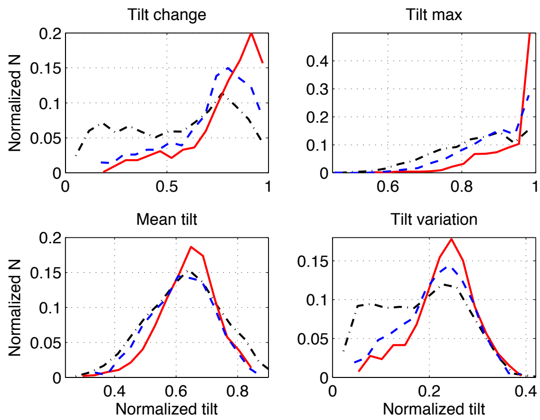

Figure 6: *Histograms with normalized frequency counts for spectral tilt. Red solid line: all stressed words, black dash-dotted line: all unstressed words, blue dashed line: unstressed without function words.*

Specifically, energy, F0, duration, and spectral tilt showed strong connections with the perception of sentential stress. In the literature, energy, F0 and duration are widely accepted as cues to stress in ADS. For instance, studies such as that of Lieberman [12] and Cutler [1] give consistent indications of their importance. In our analysis, we observed that the differences between stressed and unstressed words were particularly prominent for the majority of the features. This effect might be a result of the use of an IDS corpus. In IDS, it is typical to have an exaggeration of the features, such as increased F0 range employed by the caregiver or slower speaking rate, aimed at increasing the perceptual clarity of the communicated message (see [9]). These exaggerated changes seem to be easily perceivable by the adult listener, especially since IDS has many features commonly found in ADS clear speech (see [9]).

Finally, spectral tilt also had a clear effect in distinguishing stressed words. Currently there is no consensus on the role of tilt across languages (see, e.g., [16]) but there is good evidence of its importance as a correlate of stress (see [15]). The effect in our study may come from the increase in the employed F0 range. It was observed that, on average, the stressed words had higher F0 change in comparison to the unstressed and therefore, a corresponding change in the tilt values may have taken place.

Cross-linguistic effects in the perception of English IDS by Finnish L1 listeners may have been possible but as stress is conveyed using similar prosodic features in both languages such effect was not anticipated [32] (see also [20]). This is also evident by the observed agreement rates that were equivalent to studies carried out in American English [30,31].

In conclusion, our results show that sentence stress in IDS is conveyed in a similar way as in ADS where all the analyzed features were correlated with the perception of stress. Future work will look into expanding the existing set of data and collecting annotations from more listeners with different language backgrounds in order to model the acoustics of sentence stress across languages and examine potential differences.

5. Acknowledgements

This research was performed as a part of the Data to Intelligence (D2I) project funded by Tekes, Finland, and by the Academy of Finland in the project "Computational modeling of language acquisition". We thank all the participants in the experiment for making the study possible. The authors also thank Unto K. Laine for his valuable comments.

6. References

[1] Cutler, A., "Lexical Stress", In Pisoni, D. B. and Remez, R. E. (Eds.), *The Handbook of Speech Perception*. Malden, MA, Oxford, and Carlton, Victoria: Blackwell publishing, 264–289, 2005.

[2] Werner, S., & Keller, E., "Prosodic aspects of speech", In Keller, E. (Ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*. Chichester: John Wiley, 23–40, 1994.

[3] Wagner, P., "Great expectations – introspective vs. perceptual prominence ratings and their acoustic correlates", In *Proceedings of Interspeech*, 2381–2384, 2005.

[4] Chen, Y., Robb, M. P., Gilbert, H. R., and Lerman, J. W., "A study of sentence stress production in Mandarin speakers of American English", *Journal of the Acoustical Society of America*, 109(4), 1681–1690, 2001.

[5] Eriksson, A., Barbosa, P. A., and Akesson, J., "The acoustics of word stress in Swedish: A function of stress level, speaking style and word accent", In *Proceedings of Interspeech*, 778–782, 2013.

[6] Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastian-Galles, N., Limissuri, R. A., & Peperkamp, S., "Language-specific stress perception by 9-month-old French and Spanish infants", *Developmental Science*, 12(6), 941–919, 2009.

[7] Ostendorf, M., Price, P., and Shattuck-Hufnagel, S., "The Boston University radio news corpus", Linguistic Data Consortium.

[8] Cutler, A., & Foss, D.J., "On the role of sentence stress in sentence processing", *Language and Speech*, 20, 1–10, 1977.

[9] Song, J. Y., Demuth, K., and Morgan, J., "Effects of the acoustic properties of infant-directed speech on infant word recognition", *Journal of the Acoustical Society of America,* 128(1), 389–400, 2010.

[10] Fry, D. B., "Duration and intensity as physical correlates of linguistic stress", *Journal of the Acoustical Society of America*, 27(4), 765–768, 1955.

[11] Fry, D. B., "Experiments in the perception of stress", *Language and Speech*, 1(2), 126–152, 1958.

[12] Lieberman, P., "Some acoustic correlates of word stress in American English", *The Journal of the Acoustical Society of America*, 32(4), 451–454, 1960.

[13] Campbell, N., "Loudness, spectral tilt, and perceived prominence in dialogues", In *Proceedings of the 13th International Congress of Phonetic Sciences*, 3, 676-679, 1995.

[14] Campbell, N., & Beckman, M. E., "Stress, prominence, and spectral tilt", In Botinis, A., Kouroupetroglou, G., and Carayiannis, G. (Eds.), *Intonation: Theory, Models, and Applications (Proceedings of an ESCA Workshop)*, 67–70, 1997.

[15] Sluijter, A. M. C. & van Heuven, V. J., "Spectral balance as an acoustic correlate of linguistic stress", *Journal of the Acoustical Society of America*, 100(4), 2471–2485, 1996.

[16] Ortega-Llebaria, M., & Prieto, P., "Acoustic correlates of stress in central Catalan and Castilian Spanish", *Language and Speech*, 54(1), 1–25, 2010.

[17] Lehiste, I., & Peterson, G. E., "Vowel amplitude and phonemic stress in American English", *Journal of the Acoustical Society of America*, 31(4), 428–435, 1959.

[18] Terken, J., "Fundamental frequency and perceived prominence of accented syllables", *Journal of the Acoustical Society of America*, 89(4), 1768–1776, 1991.

[19] Kochanski, G., Grabe, E., Coleman, J., & Rosner, B., "Loudness predicts prominence: Fundamental frequency lends little", *Journal of the Acoustical Society of America*, 118(2), 1038–1054, 2005.

[20] Endress, A. D., & Hauser, M. D., "Word Segmentation with Universal Prosodic Cues", *Cognitive Psychology*, 61, 177–199, 2010.

[21] Altosaar, T., ten Bosch, L., Aimetti, G., Koniaris, C., Demuynck, K., & van den Heuvel, H, "A speech corpus for modeling language acquisition: CAREGIVER", In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 1062–1068, 2010.

[22] Zahorian, S. A., & Hu, H., "A spectral/temporal method for robust fundamental frequency tracking", *Journal of the Acoustical Society of America*, 123, 4559–4571, 2008.

[23] Minematsu, N., Kobashikawa, S., Hirose, K., & Erickson, D., "Acoustic modeling of sentence stress using differential features between syllables for English rhythm learning system development", In *Seventh International Conference on Spoken Language Processing*, 745–748, 2002.

[24] Tsiakoulis, P., Potamianos, A., and Dimitriadis, D., "Spectral moment features augmented by low order cepstral coefficients for robust ASR", *IEEE Signal Processing Letters*, 17(6), 551-554, 2010.

[25] Imoto, K., Tsubota, Y., Raux, A., Kawahara, T., and Dantsuji, M., "Modeling and automatic detection of English sentence stress for computer-assisted English prosody learning system", *Seventh International Conference on Spoken Language Processing*, 749–752, 2002.

[26] Zhang, Y., Nissen, S. L., and Francis, A. L., "Acoustic characteristics of English lexical stress produces by native Mandarin speakers", *Journal of the Acoustical Society of America*, 123(6), 4498–4513, 2008.

[27] Fleiss, J. L., "Measuring norminal scale agreement among many raters", *Psychological Bulletin,* 76, 378–382, 1971.

[28] Cohen, J. A., "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, 20, 37–46, 1960.

[29] Landis, J. R., & Koch, G. G., "The measurement of observer agreement of categorical data", *Biometrics*, 33, 159–174, 1977.

[30] You, H-J., "Determining prominence and prosodic boundaries in Korean by non-expert rapid prosody transcription", In *Proceedings of 6th Speech Prosody*, Shanghai, China, 2012.

[31] Mo, Y., Cole, J., & Lee, E-K., "Naïve listeners' prominence and boundary perception", In *Proceedings of 4th Speech Prosody*, Campinas, Brazil, 735–738, 2008.

[32] Iivonen, A., Niemi, T., & Paananen, M., "Do F0 peaks coincide with lexical stresses? A comparison of English, Finnish, and German", In *Nordic prosody: Proceedings of the VII Conference*, Joensuu, Finland, 142–158, 1996.