Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

The Acoustic Correlates of Stress and Accent in English Content and Function Words

Robert Fuchs

Englisches Seminar, Westfälische Wilhelms-Universität Münster, Germany
robert.fuchs@uni-muenster.de

Abstract

This paper has two aims: (1) To contribute to the discussion on what the acoustic correlates of stress and accent in English are, a question on which there is currently no universal agreement; (2) To determine whether vowels in function words receive less stress than similarly unstressed vowels in content words. To this purpose, the study analyses 614 occurrences of the lax high front vowel /ɪ/ in read speech produced by 10 male speakers of Standard Southern British English. 14 different acoustic features are investigated.

Results indicate that (1) there are two acoustic correlates of **accent** (duration and f_0 slope), four acoustic correlates of **stress** (spectral balance/tilt, intensity/loudness, amplitude of voicing (H1), amplitude of the first harmonic (A1), H1*-A2 and H1*-A3*), one potential acoustic correlate of **prominence in general** (F1), and four acoustic features that appear to be unrelated to the expression of accent, stress or prominence (F2, HNR, glottal leakage (B1) and the open quotient (H1*-H2*)).

Regarding question (2), there is also limited evidence that British English function words might be less prominent than unstressed syllables in content words.

Index Terms: Acoustic correlates of stress, acoustic correlates of accent, content words, function words, British English, Standard Southern British English, spectral balance, duration, intensity, loudness, f_0, harmonics-to-noise ratio

1. Introduction

Prominence in English is usually described as a four-way contrast between (from most to least prominent) syllables with a pitch accent, unaccented syllables with primary stress, syllables with secondary stress, and untressed syllables [1]. There is widespread agreement in the literature that accent and stress are realised acoustically by different, but perhaps partially overlapping, sets of acoustic correlates. What these acoustic correlates are is an important question for an empirically grounded description of the prosody of English. Moreover, it is also of significance for research on the second language acquisition of stress and accent [2–4], and for research on speech rhythm, which some recent approaches conceptualise as being based on an alternation of prominent and non-prominent syllables [5–10].

Gussenhoven (2004) [1] characterises the acoustic correlates of stress as duration, spectral balance and centralisation (of vowels), which all depend on the greater articulatory effort involved in the production of stress. Intensity is expressly excluded as an acoustic correlate of stress in this standard reference. Fundamental frequency (f_0) is also excluded, and instead related to the presence of a pitch accent.

However, this description of the acoustic correlates of the production of stress and accent is not universally supported by the available empirical evidence, even if only results for English are considered. Table 1 presents an overview of previous research, indicating which studies have found clear support (”Yes”), weak support (”(Yes)”) and no support for the identification of various acoustic correlates.

More recently, several studies have focused on acoustic correlates in the perception of stress [18, 21, 22]. However, the relationship between the production and perception of stress and accent is complex [23], and perception studies do not necessarily shed light on the acoustic correlates relevant for production.

In summary, there is no agreement on the acoustic correlates of the production of stress and accent in English. Some of the diverging results might be due to differences in methodology. For example, when disentangling the effects of stress and accent, some studies compare accented syllables with primary stress to unaccented syllables with secondary stress within the same word, while others compare accented and unaccented syllables with primary stress across words. The present study will take the latter approach, and will test the relevance of a greater number of acoustic correlates than previous research.

Another issue in the analysis of the acoustic correlates of prominence is how many levels of prominence are distinguished. While most authors operationalise prominence as a three- or four-way contrast, others regard it as a scalar phenomenon with potentially more levels [24, 25]. This is supported by research on Dutch, a language closely related to English, which provided evidence that function words are realised less prominently than similar, unstressed syllables in content words [26]. Whether function words differ in stress from unstressed syllables in content words has so far not been investigated for English. Since Dutch and English may be relatively similar in the hierarchy of acoustic correlates of stress and accent [14], it is conceivable that function words differ in prominence from unstressed syllables in content words in English.

2. Aims and Approach

The aims of this pilot study are to determine (1) the acoustic correlates of stress and accent in English and (2) whether syllables in function words differ in prominence from similar syllables in content words. To this purpose, data from a text passage read by 10 male speakers of Standard Southern British English (BrE) will be analysed. The present paper focusses on a single vowel phoneme (/ɪ/), since the magnitude of acoustic correlates of prominence may differ between phonemes [12]. The lax front high vowel /ɪ/ is ideally suited for an analysis of the acoustic correlates of prominence because it is the only vowel that occurs frequently across all levels of prominence, i.e. in accented, stressed, unstressed position and in function words. A comparison with other vowels, while useful, is beyond the

 doi: 10.21437/SpeechProsody.2016-89

	Accent			Stress		
	Yes	(Yes)	No	Yes	(Yes)	No
Duration	[11–13]	[14, 15]	[16, 17]	[12, 14, 18, 19]		[16]
Fundamental frequency (f_0)	[11, 13, 14, 18, 20]	[15]			[14]	[19]
Intensity/loudness	[11, 13–15, 18]			[19]	[14]	
Spectral balance/spectral tilt	[16]	[15]			[20]	[16]
Open Quotient (H1*-H2*)	[14]		[18]		[20]	
Amplitude of Voicing (H1*)	[14]					
Aperiodicity/HNR		[15]				
Amplitude of the first harmonic (A1)	[18]					
Glottal parameters (H1*-A2 and H1*-A3*)				[14, 18]		
Glottal leakage (B1)				[14, 18]		
Vowel formants			[18]	[1]		[18]

Table 1: Acoustic correlates of accent and stress, as identified by previous research on English. "Yes" indicates clear support, "No" no support for the identification of an acoustic correlate, and "(Yes)" that the relationship is only weak.

scope of this paper.

In these data, the following differences are expected to be found, depending on wether a given acoustic parameter is an acoustic correlate of accent, stress or prominence in general: (1) If an acoustic parameter is an acoustic correlate of **accent**, we would expect a difference between vowels in accented syllables on the one hand, and vowels in stressed unaccented syllables, unstressed syllables in content words and function words, on the other hand. (2) If an acoustic parameter is an acoustic correlate of **stress**, we would expect a difference between vowels in accented syllables and stressed unaccented syllables on the one hand, and unstressed syllables in content words and function words, on the other hand. (3) If an acoustic parameter is an acoustic correlate of **prominence** in general, it should in- or decrease from vowels in accented syllables, via stressed unaccented syllables, unstressed syllables in content words, to function words.

3. Data and Methods

3.1. Data and Elicitation Methods

The analysis relies on recordings of a text passage read by 10 male speakers of BrE from the DyViS database [27]. The speakers were between 18 and 25 years old at the time of recording (2005-2009), which took place in a sound-treated studio.

3.2. Analysis

Approximately two thirds of the reading passage (392 words) were segmented based on phonemic forced alignment with HTK [28] and P2FA [29]. All annotations were manually corrected, and annotated for the presence of a pitch accent and stress [30].

Subsequently, a Praat script (available in the digital appendix) extracted all acoustic measures for all occurrences of /ɪ/. Spectral balance was measured as the difference between the amplitude at 1-4 kHz, and 0-1 kHz, f_0 as the difference between maximum and minimum f_0 (in semitones, relative to 100 Hz) within the vowel. All glottal parameters were measured based on a long term average spectrum (ltas) across the duration of the vowel. Ltas and f_0 measurements could not be derived for 59 and 96 tokens, respectively, because the algorithm failed (mainly for vowels that were rather short or creaky-voiced). H1, H2 and A3 were corrected following [31]. All other acoustic measurements were conducted based on Praat standard parameters for male speakers, and measured at vowel midpoint unless they are mean or peak values. In total, the analysis relied on 259 phonemes occurring in function words (e.g. *in*), 273 in unstressed syllables in content words (e.g. *believed*),[1] 45 in stressed unaccented syllables (e.g. *city*) and 37 in accented syllables.

Next, mixed effects regression models were run in R with LME4 [32, 33]. Each of the acoustic correlates was used in turn as the dependent variable in a regression model, with STRESS (accented, stressed, unstressed, function) as dependent variable, and SPEAKER, WORD/PRECEDING PHONEME, and WORD/FOLLOWING PHONEME as random factors. For some of the regression analyses, candidate models did not converge, so that one random variable had to be removed. To ensure that conditions for regression models are met, data points 2.5 standard deviations below or above the mean were then trimmed with function romr.fnc from the package LMERCONVENIENCEFUNCTIONS (never more than 2.5% of the data) [34]. Finally, post-hoc Tukey tests with alpha-level corrected for multiple comparisons were conducted with the glht function from package MULTCOMP [35].

4. Results

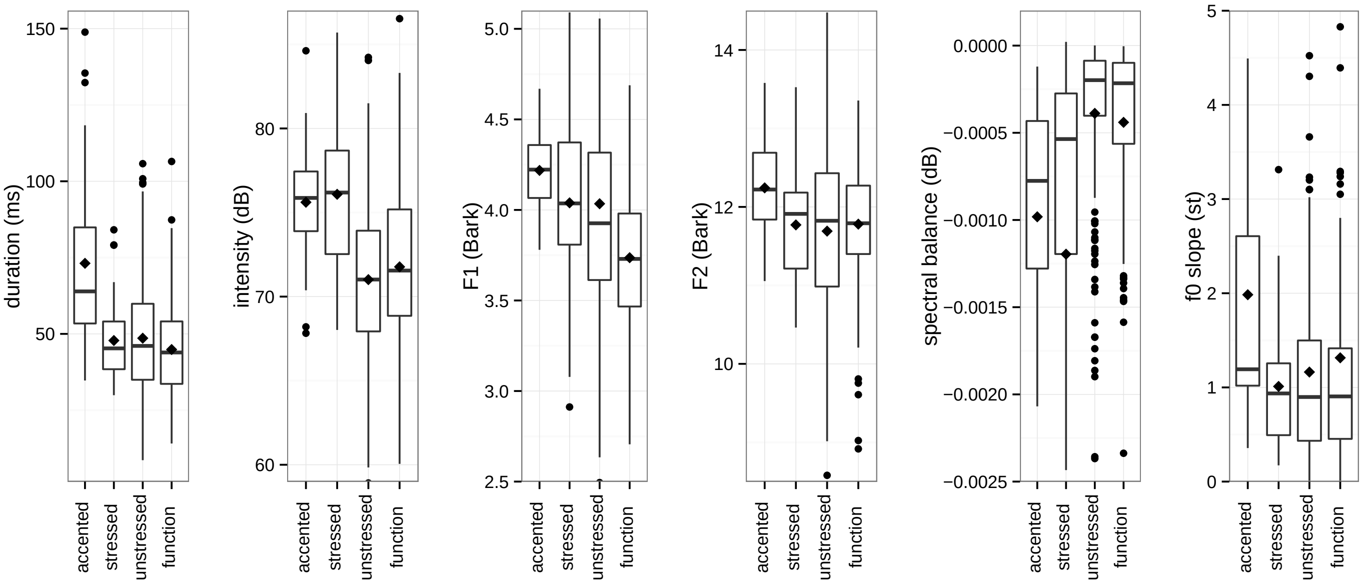
4.1. Duration

With a mean duration of 73.1 ms, vowels in accented syllables are significantly longer than stressed vowels (47.8 ms, $z=3.6$, $p<0.01$, see Fig. 1a), unstressed vowels in content words (48.6 ms, $z=4.1$, $p<0.001$), and vowels in function words (44.9 ms, $z=4.3$, $p<0.001$). All other differences are not significant. The results suggest that duration is an acoustic correlate (AC) of accent, but not of stress.

4.2. Intensity and Loudness

Vowels in accented and in unaccented stressed syllables in content words do not differ in peak intensity (75.6 dB, 76.1 dB, $z=0.02$, $p=1.0$, see Fig. 1b). They are louder than unstressed vowels in content words (71.0 dB, $z=3.3$, $p<0.01$; $z=3.4$, $p<0.01$), and vowels in function words (71.8 dB, $z=2.9$, $p<0.05$; $z=3.0$, $p<0.05$). Unstressed vowels in content and

[1]The manual annotation showed that none of the canonically unstressed vowels in content or in function words were accented.



(a) duration (b) intensity (c) F1 (d) F2 (e) spec. balance (f) f0 slope

Figure 1: Acoustic correlates of prominence in vowels in *accented* syllables, *stressed* unaccented syllables, *unstressed* syllables in content words and in *function* words (some outliers not shown). Diamonds indicate means, horizontal lines medians, boxes extend to the 25th and 75th percentiles, and whiskers to the highest/lowest point from the box that is within 1.5 times the interquartile range.

function words do not differ in peak intensity (z=0.6, n.s.). Results for peak and mean loudness are similar, but in some cases reach better significance levels. Results for mean intensity are also relatively similar, with the exception that the difference between vowels in stressed unaccented syllables and function words is only borderline significant (z=-2.4, p<0.08).

4.3. First and Second Formants (F1 and F2)

There is a small difference in F1 between vowels in accented syllables (4.2 Bark) and vowels in stressed and unstressed syllables in content words (both 4.0 Bark, z=0.9, z=0.5, n.s., see Fig. 1c). Vowels in function words in turn have a somewhat lower F1, but none of the differences are significant (3.7 Bark, z=1.3, z=1.4, z=2.2, n.s.). When stress level is quantified on a numerical scale from 4 (accented) to 1 (function words), there is a weak but highly significant correlation between stress level and F1 (Spearman's ρ=0.31, p<0.001). This suggests that F1 may be an AC of prominence in general, but not specifically of stress or accent.

For F2, there are also small, but no significant differences between vowels in accented syllables, on the one hand (12.2 Bark), and vowels in stressed and unstressed syllables (11.8 Bark, 11.7 Bark), and in function words, on the other hand (11.8 Bark; z between 0.2 and 1.0, n.s., see Fig. 1d). The results suggest that F2 is not sensitive to stress or accent.

4.4. Spectral Balance

While accented and stressed syllables do not differ in spectral balance (-0.00098 dB, -0.00120 dB, z=0.9, n.s., see Fig. 1e), they both have a more skewed spectrum than unstressed syllables in content (-0.00039 dB, z=3.7, p<0.01, z=2.8, p<0.05) and function words (-0.00044 dB, z=3.3, p<0.01, 2.5, p<0.06). Unstressed vowels in content and function words do not differ in spectral balance (z=0.5, p>0.95). This suggests that spectral balance is an AC of stress.

4.5. Fundamental Frequency (f_0)

The analysis of f_0 slope shows that vowels in accented syllables have a signficantly steeper slope (2.0 semitones (st)) than vowels in stressed syllables (1.0 st, z=3.5, p<0.01), in unstressed syllables in content words (1.2 st, z=3.3, p<0.01) and in function words (1.3 st, z=3.1, p=0.01, see Fig. 1f). All other differences are not significant. In particular, vowels in unstressed syllables in content words do not differ from vowels in function words (z=0.1, p>0.99). These results suggests that f_0 slope is an AC of accent.

4.6. Harmonics-to-Noise Ratio (HNR)

Vowels in accented syllables (mean 19.7 dB) have a higher peak HNR than vowels in stressed unaccented syllables (17.8 dB), unstressed syllables in content words (18.3 dB) and function words (18.3 dB), but none of the differences are significant (see Fig. 2a). Results for HNR at vowel midpoint are similar and therefore not described in detail. These results suggest that HNR is not sensitive to stress or accent.

4.7. Glottal Leakage/Bandwidth of F1 (B1)

Vowels in accented syllables (mean 0.70 Bark) and in stressed unaccented syllables (0.72 Bark) have a somewhat lower B1 than vowels in unstressed syllables in content words (0.91 Bark) and function words (0.82 Bark), but none of the differences are significant (see Fig. 2b). This suggests that B1 is not sensitive to stress or accent.

4.8. Glottal Parameters (H1*-A2 and H1*-A3*)

Closure rate of the glottal pulse (H1*-A2) is lowest for vowels in accented syllables (10.8 dB), and slightly but insignifi-

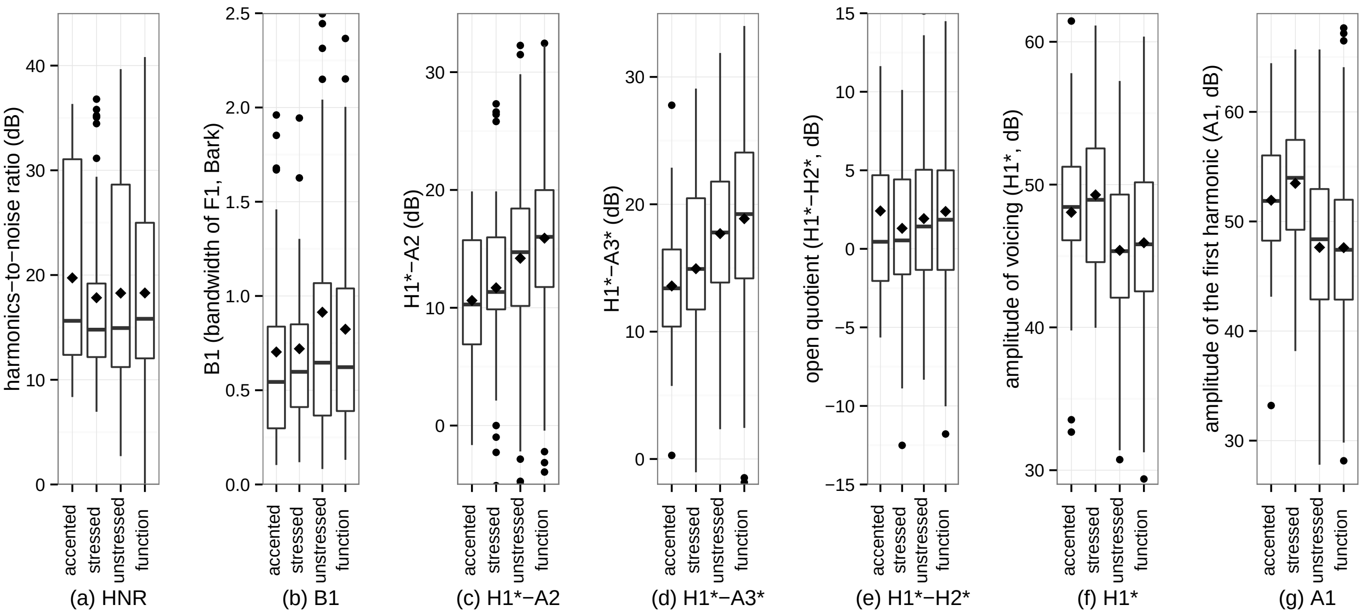


Figure 2: Acoustic correlates of prominence in vowels, see Fig. 1 for details.

cantly higher in stressed unaccented syllables (11.7 dB, z=0.1, p>0.999, see Fig. 2c). In unstressed syllables in content words, it is still higher (14.1 dB), but not significantly different from stressed unaccented syllables (z=1.9, n.s.). Vowels in function words have the highest H1*-A2 (15.9 dB), which does not differ significantly from unstressed syllables in content words (z=1.4, n.s.), but differs significantly from stressed unaccented and accented syllables (z=3.0, z=2.0, p<0.05).

Results for skewness of the glottal pulse (H1*-A3*) are comparable, in that it is lowest for vowels in accented syllables (13.6 dB), followed by stressed unaccented syllables (14.9 dB), unstressed syllables in content words (17.7 dB), and vowels in function words (18.7 dB, see Fig. 2d). Only the differences between function words, on the one hand, and accented and stressed unaccented syllables, on the other hand, are significant (z=2.9; z=2.6, p<0.05). These results suggest that (1) H1*-A2 and H1*-A3* are ACs of stress and (2) unstressed syllables in function words might be less prominent than unstressed syllables in content words in the sense that only the former differed signficantly from accented and stressed syllables.

4.9. Open Quotient (H1*-H2*)

Results for H1*-H2* reveal no significant differences between any of the conditions (see Fig. 2e). This suggests that H1*-H2* may not be an AC of stress or accent.

4.10. Amplitude of Voicing (H1*) and the First Harm. (A1)

H1* is higher for vowels in accented and stressed unaccented syllables (47.8 dB, 49.3 dB) than in vowels in unstressed syllables in content words and function words, on the other hand (45.5 dB, 46.0 dB, see Fig. 2f). There is a significant difference between stressed unaccented syllables and unstressed syllables in content words (z=2.6, p<0.05), and a borderline significant difference between stressed unaccented syllables and function words (z=2.5, p=0.05).

A1 is higher for vowels in accented and stressed unaccented syllables, on the one hand (51.9 dB, 53.5 dB), than for vowels in unstressed syllables in content words and function words, on the other hand (47.6 dB, 47.6 dB, see Fig. 2g). There is a significant difference between vowels in stressed unaccented syllables, on the one hand, and unstressed syllables in content words and function words, on the other hand (z=2.9, p<0.05; z=2.7, p<0.05). This suggests that H1* and A1 are ACs of stress.

5. Discussion and Conclusion

This paper investigated acoustic correlates of stress and accent in 614 occurences of the vowel /ɪ/ across four levels of prominence (accented, stressed unaccented, unstressed in content words, unstressed in function words) in Standard Southern British English. Results indicate that

- duration and f_0 slope are acoustic correlates of **accent**
- spectral balance/tilt, intensity/loudness, amplitude of voicing (H1*) and the first harmonic (A1), as well as the glottal parameters H1*-A2 and H1*-A3* are acoustic correlates of **stress**
- F1 may be an acoustic correlate of **prominence** in general
- F2, HNR, glottal leakage (B1) and the open quotient (H1*-H2*) are not related to the expression of accent, stress or prominence.

Regarding the question of whether syllables in function words are less prominent than unstressed syllables in content words, there was no evidence of significant differences between these two levels in any of the acoustic correlates. However, the results for several acoustic correlates (F1, H1*-A2, H1*-A3*) indicated that vowels in function words are slightly less prominent than unstressed vowels in function words. Future research will have to show whether this and the other findings can be confirmed by analyses of greater numbers of speakers and other vowels.

6. References

[1] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[2] J. Li and P. P. Mok, "The acquisition of english lexical stress by cantonese-english bilingual children at 2;06 and 3;0," in *Proceedings of Speech Prosody 7*, N. Campbell, D. Gibbon, and D. Hirst, Eds., Dublin, 2014, pp. 688–692.

[3] Y. Zhang, S. L. Nissen, and A. L. Francis, "Acoustic characteristics of english lexical stress produced by native mandarin speakers," *Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4498–4513, 2008.

[4] M.-C. Michaux, S. Brognaux, and G. Christodoulides, "The production and perception of l1 and l2 dutch stress," in *Proceedings of Speech Prosody 7*, N. Campbell, D. Gibbon, and D. Hirst, Eds., Dublin, 2014, pp. 462–466.

[5] R. Fuchs, *Speech Rhythm in Varieties of English. Evidence from Educated Indian English and British English*. Singapore: Springer, 2016.

[6] ——, "Integrating variability in loudness and duration in a multidimensional model of speech rhythm: Evidence from indian english and british english," in *Proceedings of Speech Prosody 7, Dublin*, N. Campbell, D. Gibbon, and D. Hirst, Eds., 2014, pp. 290–294.

[7] ——, "Towards a perceptual model of speech rhythm: Integrating the influence of f0 on perceived duration," in *Proceedings of Interspeech 2014*, H. Li, H. Meng, B. Ma, E. S. Chng, and L. Xie, Eds., Singapore, 2014, pp. 1949–1953.

[8] L. He, "Syllabic intensity variations as quantification of speech rhythm: Evidence from both l1 and l2," in *Proceedings of the 6th International Conference on Speech Prosody, Shanghai, May 22-26, 2012*, Q. Ma, H. Ding, and D. Hirst, Eds. Shanghai: Tongji University Press, 2012, pp. 466–469.

[9] R. E. Cumming, "The language-specific integration of pitch and duration," Ph.D. dissertation, University of Cambridge, 2010.

[10] A. Galves, J. Garcia, D. Duarte, and C. Galves, "Sonority as a basis for rhythmic class discrimination," in *Proceedings of Speech Prosody 2002*, 2002, pp. 323–326.

[11] M. E. Beckman, *Stress and Non-Stress Accent*. Dordrecht/Riverton: Foris, 1986.

[12] K. de Jong, "Stress, lexical focus, and segmental focus in english: Patterns of variation in vowel duration," *Journal of Phonetics*, vol. 32, pp. 493–516, 2004.

[13] P. R. Keyworth, "The acoustic correlates of stress-shifting suffices in native and nonnative english," Master's thesis, St. Cloud State University, 2014.

[14] A. Sluijter and V. van Heuven, "Acoustic correlates of linguistic stress and accent in dutch and american english," in *Proceedings of the International Congress of Spoken Language Processing*, Philadelphia, 1996, pp. 630–633.

[15] G. Kochanski, E. Grabe, J. Coleman, and B. Rosner, "Loudness predicts prominence: Fundamental frequency lends little," *Journal of the Acoustical Society of America*, vol. 118, p. 2005, 2005.

[16] N. Campbell and M. E. Beckman, "Stress, prominence, and spectral tilt," in *Intonation: Theory, Models and Applications*, A. Botinis, G. Kouroupetroglou, and G. Carayiannis, Eds. Athens: ESCA and University of Athens Department of Informatics, 1997, pp. 67–70.

[17] M. Beckman and J. Edwards, "Articulatory evidence for differentiating stress categories," in *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, P. A. Keating, Ed. Cambridge: Cambridge University Press, 1994, pp. 7–33.

[18] A. O. Okobi, "Acoustic correlates of word stress in american english," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.

[19] V. Huss, "English word stress in the post-nuclear position," *Phonetica*, vol. 35, pp. 86–105, 1978.

[20] Y.-L. Shue, M. Iseli, N. Veilleux, and A. Alwan, "Pitch accent versus lexical stress: quantifying acoustic measures related to the voice source," in *Proceedings of Interspeech 2007*, 2007, pp. 2625–2628.

[21] D. Arnold, B. Möbius, and P. Wagner, "Comparing word and syllable prominence rated by naive listeners," *Proceedings of Interspeech 2011*, 2011.

[22] A. Chrabaszcz, M. Winn, C. Y. Lin, and W. J. Idsardi, "Acoustic cues to perception of word stress by english, mandarin, and russian speakers," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 4, pp. 1468–1479, 2014.

[23] V. J. van Heuven, "Acoustic correlates and perceptual cues of word and sentence stress: Mainly english and dutch," in *4th International Symposium on Tonal Aspects of Languages, May 13-16 2014*, Nijmegen, 2014.

[24] M. Liberman and A. Prince, "On stress and linguistic rhythm," *Linguistic Inquiry*, vol. 8, pp. 249–336, 1977.

[25] M. Halle and J.-R. Vergnaud, *An Essay on Stress*. Cambridge: MIT press, 1987.

[26] D. R. Van Bergem, "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," *Speech Communication*, vol. 12, no. 1, pp. 1–23, 1993.

[27] F. Nolan, K. McDougall, G. de Jong, and T. Hudson, "A forensic phonetic study of dynamic sources of variability in speech: The dyvis project," in *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, P. Warren and C. Watson, Eds., 2006, pp. 13–18.

[28] S. J. Young, *The HTK Hidden Markov Model Toolkit: Design and Philosophy*, University of Cambridge: Department of Engineering Std., 1993.

[29] J. Yuan and M. Liberman, "Speaker identification on the scotus corpus," in *Proceedings of Acoustics '08*, 2008, pp. 5687–5690.

[30] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer (Computer Program). Version 5.3.04*, Std. [Online]. Available: www.praat.org

[31] H. M. Hanson, "Glottal characateristics of female speakers: Acoustic correlates," *Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 466–481, 1997.

[32] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008. [Online]. Available: http://www.R-project.org

[33] D. Bates, M. Maechler, and B. Bolker, *lme4: Linear Mixed-Effects Models Using S4 Classes*, 2013, r package version 0.999999-2. [Online]. Available: http://CRAN.R-project.org/package=lme4

[34] A. Tremblay, D. University, J. Ransijn, and U. of Copenhagen, *LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models*, 2015, r package version 2.10.

[35] T. Hothorn, F. Bretz, and P. Westfall, "Simultaneous inference in general parametric models," *Biometrical Journal*, vol. 50, no. 3, pp. 346–363, 2008.