Stress, Prominence, and Spectral Tilt

Nick Campbell and Mary Beckman †
ATR Interpreting Telecommunications Research Laboratories
†Ohio-State University.
nick@itl.atr.co.jp, mbeckman@ling.ohio-state.edu

Abstract

This paper examines spectral correlates of stress and accent in a corpus of sentences with varying focus, produced by four speakers of American English. Analyses of the spectrum at vowel centre show a clear effect on the spectral tilt – i.e. more energy at higher frequencies relative to energy nearer to the fundamental – when the vowel is in a nuclear-accented syllable. However, unlike in the Dutch corpus examined by Sluijter & van Heuven, there was no difference in spectral tilt between vowels in stressed versus unstressed syllables in the absence of the accompanying intonational prominence contrast. These results lend support to the hypothesis that syllables marking focal prominence are phonated in a more emphatic way than other syllables. That is, accented syllables may be louder and not simply intonationally prominent, but this effect does not distinguish an independent lexically specified level of 'primary stress' between the intonational prominence of accent and the basic rhythmic contrast between strong (full) and weak (reduced) syllables.

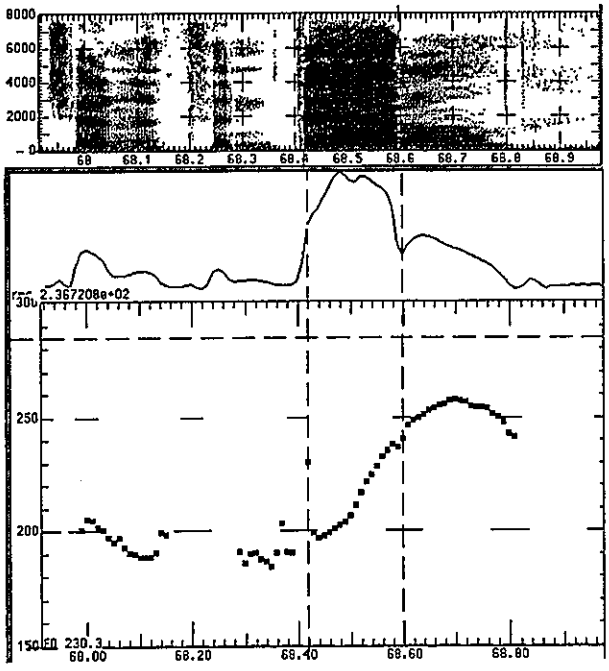

Figure 1: Spectrogram, RMS intensity, and F0 for the acoustic sequence 'Jonathan Baddle' in pattern 1 where target syllable (delimited with cursors) is the accented syllable (underlined)

1 Introduction

In English, Dutch, and other 'stress-accent languages', pitch accents typically occur only on lexically prominent syllables. For example, in the name *Baddle*, only the strong first syllable can be accented, as it is in the intonation contour displayed in Fig. 1. The weak (reduced) second syllable does not bear a pitch accent, and cannot bear the nuclear pitch accent except in very marked contexts, such as metalinguistic correction of pronunciation. While the two syllables of *Baddle* might suggest otherwise, this constraint on accent location is over and above the basic level of contrast between strong (full) and weak (reduced) syllables. This can be appreciated by comparing the name *Badd-Ellis* in the same intonational context. Both names have strong first syllables, which contrast with the weak second syllable of *Baddle* and the weak third syllable of *Badd-Ellis*. But the location of nuclear accent would differ; in *Badd-Ellis* it would fall on the second syllable.

This difference in accentability corresponds roughly to the traditional distinction 'primary stress' and 'secondary stress'. However, a syllable with 'primary stress' is not accented in all discourse contexts. For example, in the utterances in Figs. 2 and 3, the nuclear pitch accent falls on the first syllable of the given name *Jonathan* instead of on the first syllable in *Baddle* or the second syllable in *Badd-Ellis*. Structural descriptions, therefore, must differentiate 'stress' proper (i.e. the lexical marking of the accentable syllable) from 'accent' (i.e. the intonational marking of the syllable with the 'nuclear stress'). Two fundamental questions, therefore, are whether there are reliable phonetic correlates of this structural difference in English, and if so, what these correlates might be.

Traditional accounts described 'stress' as a local increase in loudness, differentiating it thus from the intonational event of the pitch accent. Experiments such as Fry (1955, 1958), suggest that the traditional account is not tenable. These experiments show that overall RMS intensity can differentiate

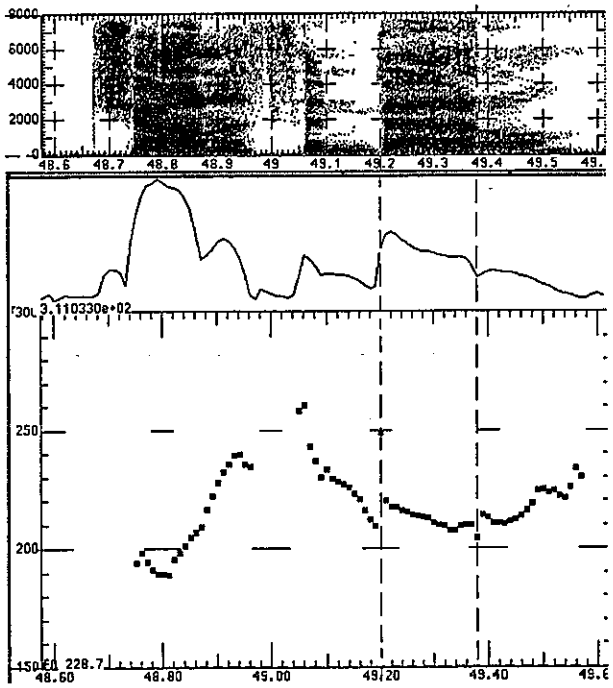

Figure 2: Spectrogram, RMS intensity, and F0 for 'Jonathan Baddle' in pattern 2 where target syllable is after the accented syllable

stressed from unstressed syllables, but this measure is not a reliable indicator of the structural difference in the absence of rigid control for other things that affect RMS intensity – including the fundamental frequency. Moreover, varying overall RMS intensity to match the acoustic differences has very little effect on the percept of stress – particularly by comparison to the perceptually very salient effects of manipulat-

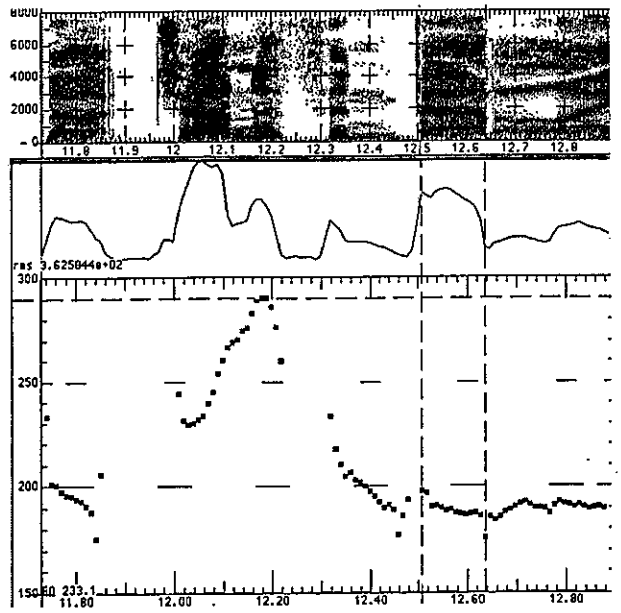

Figure 3: Spectrogram, RMS intensity, and F0 for 'Jonathan Baddle' in pattern 3 where target syllable is after the accented one

ing the fundamental frequency to make even the very roughest approximations to intonation patterns that can be parsed as placing nuclear accent on the target vowel.

However, Sluijter and van Heuven (1996) point out that Fry's studies and all subsequent similar studies co-vary accent and stress. They suggest that loudness can be shown to be more a correlate of lexical stress if a more appropriate measure of loudness is chosen than overall intensity. They examined the balance of energy in higher spectral bands relative to lower frequencies in a Dutch corpus that put minimally contrasting words in focal and in non-focal position, and found an effect of primary stress even outside of focal accent position. We extended their work to examine similar acoustic correlates in target vowels in a corpus of English materials that varies 'lexical stress' independently of accentual prominence.

2 Materials

We controlled for three types of vowels ([ae], [i], [u]) in 2 types of initial syllables:

```
1. Baddle, Beedle, Boodle (full primary lexical stress)
2. Badd-Ellis, Beede-Ellis, Boode-Ellis
                          (unreduced but unstressed)
```

placed in 3 intonational contexts:

1. syllable with *High** nuclear pitch accent (Fig. 1):

```
He's met ALL of the men from that gang.  He's met
Tony LUCIANO, Jonathan BADDLE, Nathaniel JACKSON...
                    (L+)H* H-
```

2. postnuclear unaccented syllable in *High* context (Fig. 2):

```
He's written books on ALL of the famous Baddles. He's
done MATTHEW Baddle, JONATHAN Baddle, MIRIAM Baddle, ...
                    H*        H-
```

3. postnuclear unaccented syllable in *Low* context (Fig. 3):

```
No, it's not JONATHAN Baddle I interviewed, but his
                 H*       L-
brother, Matthew.
```

produced 10 times each by 3 female speakers (JV, MB [the second author], MO) and 1 male speaker (KL) of American English. Recordings were made in a noiseless acoustically treated chamber.

Table 1: Corpus materials provided for the following comparisons to be made :

Accent:	accented	unaccented	
Tone:	(L+)H*	H-	L-
stressed	1-1	2-1	3-1
unstressed		2-2	3-2

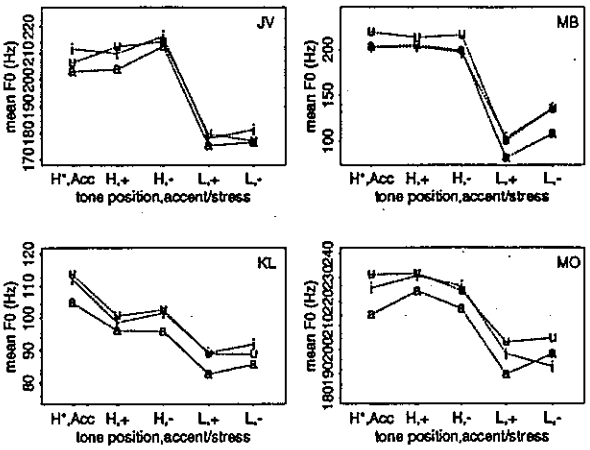

Figure 4: Mean values for the average vowel F0 in the 5 comparison contexts.

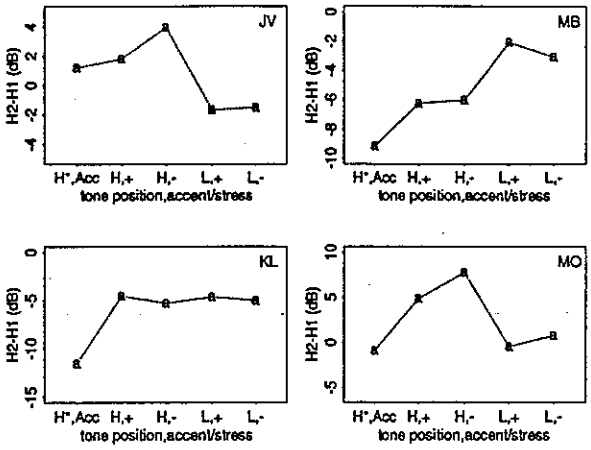

Figure 6: Mean values for the intensity ratio between second and first harmonics.

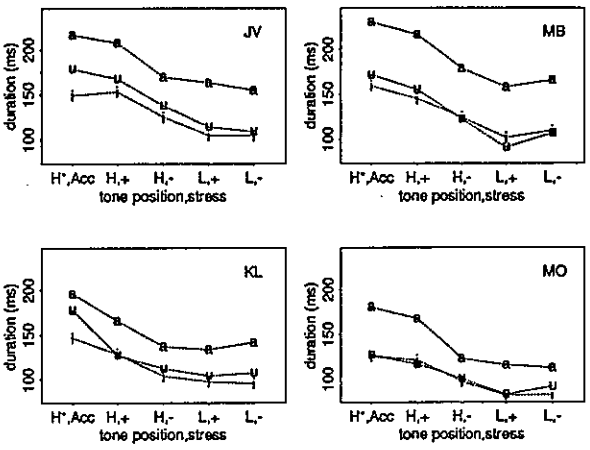

Figure 5: Means for vowel duration in the 5 comparison contexts.

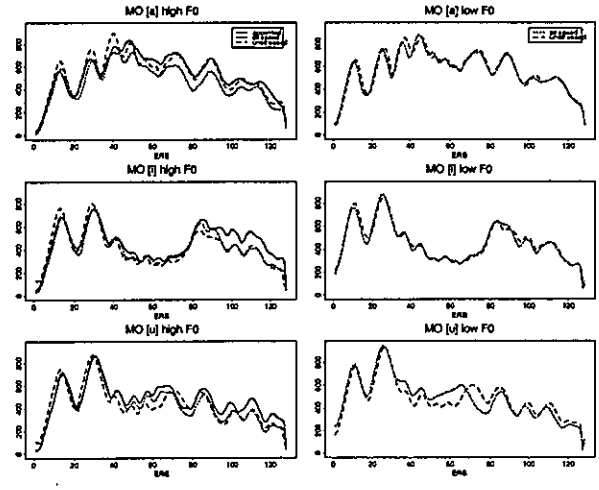

Figure 7: ERB power spectrum for speaker MO. (KEY: solid line: accented, dotted line: stressed, dashed line: unstressed).

- **1-1 vs 2-1:**
  Accented syllable to unaccented-but-stressed syllable (Baddle, etc.) in *H-* tone context (F0 rises to *H** peak in 1-1, but value at V center, and mean F0 over V are the same as *H-* level in 2-2).

- **2-1 vs 2-2:**
  Unaccented-but-stressed (Baddle) to unstressed (Badd-Ellis) in postnuclear *H-* tone context (F0 is level and high in target V in both contexts).

- **3-1 vs 3-2:**
  Unaccented-but-stressed (Baddle) to unstressed (Badd-Ellis) in postnuclear *L-* tone context (F0 trace for unstressed counterpart is identical).

2.1 Acoustic measures

After recording and digitization, we excised the 720 target vowel portions of the speech and measured their duration, fundamental frequency (F_0) and RMS amplitude from the waveform. We calculated the mean fundamental frequency for every excised vowel token (F0 in Hz). We measured the length of the target vowel interval (duration in ms). We computed vowel formant frequencies at 10 ms frames and calculated average frequencies over 3 frames at center of vowel (F1, F2 in Hz). We extracted amplitudes of harmonics averaged over a 30 ms window at vowel center and calculated an intensity ratio between first and second harmonics (H2-H1 in dB), a standard measure of spectral tilt in phonetic studies of contrastive phonation type (see Jackson 1985). We computed a 128-channel ERB-scaled filterbank analysis for successive 10 ms frames through the vowel (after Patterson 1986) and calculated an average power spectrum (ERB) for the center 3 frames of each token.

3 Results

Fig. 4 shows mean values for the average F0 measure. The four talkers were consistent in producing the intended intonation patterns, except that KL produced a level H* accent rather than a rising L+H* accent in intonation type 1-1, making the mean F0 values for the vowel in this type higher than for the unaccented vowel in the comparison between accented (1-1) and unaccented/stressed (2-1).

Fig. 5 shows mean values for the duration measure. The four talkers varied in how reliably they differentiated accented from unaccented syllables by this measure of prominence. They differentiated stressed from unstressed syllables only in the H- tone context.

Fig. 6 shows mean values for the intensity ratio between second and first harmonics, which differentiated accented target vowels from unaccented vowels in some contexts. However, this was not a reliable correlate of either level of prominence contrast because it was very sensitive to other uses of phonation type contrasts. In particular, two talkers had creaky-voice phonation as a correlate of the percept of the low tone in intonation pattern 3.

Fig. 7 shows ERB spectra, averaged over the tokens of each of the five comparison types for speaker MO. For all three vowels, the accented tokens are differentiated from the unaccented-but-stressed tokens in having considerably higher energy levels just in the higher frequency bands (ERB 60 and above). Stressed and unstressed tokens are not differentiated by this pattern. (In the High tone context, energy is higher in the unstressed [ae] in these bands, but not relatively higher compared with the lower bands.) The other three talkers showed the same pattern, albeit with a less striking difference, and somewhat less consistency across the three vowel types.

4 Discussion and conclusion

We found a difference between accented and unaccented syllables, replicating Sluijter & van Heuven's (1996) findings for spectral balance differences in Dutch between vowels produced with and without focal accent. However, unlike Sluijter & van Heuven, we found virtually no difference for this measure between stressed versus unstressed syllables in the absence of the accent contrast. Other correlates of 'stress' proper were variable across speaker and vowels, and interacted with the pitch range defined by the intonation pattern. This failure to replicate Sluijter & van Heuven's findings is in keeping with differences between the two languages in the perceptibility of stress in the absence of accent or vowel reduction contrasts, e.g. van Heuven (1987) vs Huss (1977). Dutch differs from English in having relatively fewer words in which unstressed syllables are reduced, particularly in word initial position (Booij, 1995).

These results contradict traditional descriptions of stress as a cross-linguistic phonetic property separable from accent, and lend support to Bolinger's (1958) claim that stress in English is nothing more than a structural marking of potential for a syllable to bear pitch accent.

An important application of this type of knowledge is in the selection of speech segments for concatenative speech synthesis (Campbell 1992). Our present findings show that the commonly used prosodic modifications (duration and fundamental frequency stretching) will be insufficient to model the prominence characteristics of nuclear accents in English speech. Since current technology is unable to modify spectral characteristics without noticeable degradation of speech quality, we conclude that a larger inventory of source units will be required if we are to accurately replicate the characteristics of nuclear accented syllables to express focus in the synthesised utterances.

References

[1] D. Bolinger "A theory of pitch accent in English", Word 14, 109- 149. 1958.

[2] G. E. Booij *The Phonology of Dutch*, Clarendon Press, 1995

[3] W. N. Campbell: "Synthesis Units for Natural English Speech", Transactions of the Institute of Electronics, Information and Communication Engineers, SP 91-129, pp 55 - 62. 1992.

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

[5] D. B. Fry, "Experiments in the perception of stress", Language and Speech 1, 126-152, 1958.

[6] J. van Heuven (1987) "Stress patterns in Dutch (compound) adjectives: Acoustic measurements and perception data", Phonetica 44, 1-12, 1987.

[7] V. Huss, "English word stress in post-nuclear position", Phonetica 35, 86-105, 1978

[8] M. Jackson, P. Ladefoged, M. K. Huffman, & N. Antoñanzas-Barroso, "Measures of spectral tilt", UCLA Working Papers in Phonetics, 61, 72-8, 1985.

[9] R. D. Patterson & B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution", In B. C. J. Moore, ed., *Frequency Selectivity in Hearing*, pp. 123-177, 1986.

[10] A. Sluijter & V. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress", J. Acoust. Soc. Am. 100, 2471-2485, 1996.

[11] A. Sluijter & V. van Heuven, & J. J. A. Pacilly, "Spectral balance as a cue in the perception of linguistic stress", J. Acoust. Soc. Am. 101, 503-513, 1997.