Available online at www.sciencedirect.com
ScienceDirect
Speech Communication xxx (2009) xxx–xxx
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

On the acoustic correlates of high and low nuclear pitch accents in American English

Yen-Liang Shue [a,*], Stefanie Shattuck-Hufnagel [b], Markus Iseli [a], Sun-Ah Jun [c], Nanette Veilleux [d], Abeer Alwan [a]

[a] Department of Electrical Engineering, University of California, Los Angeles, CA 90095, USA
[b] Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[c] Department of Linguistics, University of California, Los Angeles, CA 90095, USA
[d] Department of Computer Science, Simmons College, Boston, MA 02115, USA

Received 4 June 2009; received in revised form 25 August 2009; accepted 25 August 2009

Abstract

Earlier findings in Shue et al. (2007, 2008) raised questions about the alignment of nuclear pitch accents in American English, which are addressed here by eliciting both high and low pitch accents in two different target words in several different positions in a single-phrase utterance (early, late but not final, and final) from 20 speakers (10 male, 10 female). Results show that the F_0 peak associated with a high nuclear pitch accent is systematically displaced to an earlier point in the target word if that word is final in the phrase and thus bears the boundary-related tones as well. This effect of tonal crowding holds across speakers, genders and target words, but was not observed for low accents, adding to the growing evidence that low targets behave differently from highs. Analysis of energy shows that, across target words and genders, the average energy level of a target word is greatest at the start of an utterance and decreases with increasing proximity to the utterance boundary. Duration measures confirms the findings of existing literature on main-stress-syllable lengthening, final syllable lengthening, and lengthening associated with pitch accents, and reveals that final syllable lengthening is further enhanced if the final word also carries a pitch accent. Individual speaker analyses found that while most speakers conformed to the general trends for pitch movements there were 2/10 male and 1/10 female speakers who did not. These results show the importance of taking into account prosodic contexts and speaker variability when interpreting correlates to prosodic events such as pitch accents.

Keywords: Pitch-accent correlates; Prosody; Voice quality; Tonal crowding

1. Introduction

Prosody describes the properties of speech such as rhythm, timing, intonation and stress.[1] In American English, as in many other languages, an essential role of

* Corresponding author. Address: Department of Electrical Engineering, University of California, Los Angeles, 56-125B Engineering IV Building, Box 951594, Los Angeles, CA 90095, USA. Tel.: +1 310 8252177; fax: +1 310 2672589.
E-mail addresses: yshue@ee.ucla.edu (Y.-L. Shue), stef@speech.mit.edu (S. Shattuck-Hufnagel), iseli@ee.ucla.edu (M. Iseli), jun@humnet.ucla.edu (S.-A. Jun), nanette.veilleux@simmons.edu (N. Veilleux), alwan@ee.ucla.edu (A. Alwan).

[1] Docherty (1990) defines it in the following way: "Prosody or the melody of speech is the process used to alter the meaning (linguistic prosody) or emotional force (affective prosody) of a sentence. The components of prosody are rhythm, pitch, tone and stress and they are articulated by modulation of the acoustic correlates of prosody; frequency, duration and amplitude".

0167-6393/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2009.08.005

linguistic prosody is to signal phrase-level prominence and phrasing, using tonal targets and other cues, such as duration (see Shattuck-Hufnagel and Turk (1996) for a review). In the Autosegmental-Metrical (AM) approach to intonation (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Ladd, 1996/2008), prominence is usually marked by a high or low pitch accent on the target word; phrasing is marked by a high or low boundary tone on the lengthened final syllable of the phrase; and an additional tonal element, a high or low phrase accent, controls the F_0 between the last pitch accent of a phrase and the boundary tone on the final syllable. The AM approach proposes a sparse string of such high and low tonal targets, represented independently from the segmental/word string but associated with it, to define each intonational phrase.

A challenging issue in spoken prosody is the difficulty of specifying the acoustic correlates of these tonal elements. One problem relates to the height of the target in the F_0 space: these entities are usually defined in relational terms which are difficult to quantify absolutely. For example, the F_0 level associated with a high pitch accent for one speaker, while higher than a low pitch accent for the same speaker in the same context, might correspond to the F_0 level of a low accent for another speaker. Similarly, because the overall pitch range often declines during an utterance, a paragraph or a conversational turn (Chafe, 1993; Hirschberg and Pierrehumbert, 1986), a high pitch accent late in a constituent may actually have a lower F_0 than a low pitch accent that occurred earlier. Thus, it is difficult to specify a threshold for F_0, above which the target is a high pitch accent and below which the target is a low pitch accent, for all speakers or even for a single speaker. Another problem relates to the alignment of the target with the text in time. For example, there is good evidence that the alignment of the F_0 turning point associated with one tonal target can be influenced by the position and type of adjacent tonal targets (Silverman and Pierrehumbert, 1990; Arvaniti et al., 2006; Arvaniti and Garding, 2007). Finally, in addition to questions about F_0 levels and text alignment, there are gaping holes in our understanding of other candidate acoustic correlates of tonal targets, although valuable work has been done on the parameters of intensity (e.g. Kochanski et al., 2005), duration (e.g. Turk and colleagues) and voice quality (e.g. Pierrehumbert and Talkin, 1991; Dilley et al., 1996). In this paper, we examine the acoustic correlates of two kinds of American English pitch accents, high (denoted by H^*) and low (denoted by L^*), and how the presence of adjacent phrase accents and boundary tones on the same word can affect these correlates.

This study was inspired by some puzzling results in our earlier work (Shue et al., 2007), i.e. a striking difference in the alignment of the F_0 peak associated with a H^*, in two different words which appeared in two different positions in the intonational phrase *Dagada gave Bobby doodads*. That is, when the H^* fell on the earlier target word, *Dagada*, its peak aligned near the end of its stressed syllable (-ga-), but when the H^* fell on the later word, *doodads*, it aligned earlier in its stressed syllable (doo-). Because the two pitch accent contexts differed in several ways, it was not possible to determine which of several potential factors was responsible for the difference in peak alignment. For example, the two words differed in the quality of the stressed vowel, /a/ vs /u/, a factor that has been proposed by Jilka and Möbius (2007) to influence alignment. In that work, statistical analysis showed a correlation between vowel height and peak alignment, with high-vowel and low-vowel peak positions differing by approximately 11%. However, the corpus consisted of uncontrolled sentences selected from a newspaper corpus, which made it difficult to exclude other possible influences such as phrase position and stress as possible contributors to those results. Shue et al. (2007)'s target words *Dagada* and *doodads* also differed in a number of other ways, which may have contributed to the observed alignment difference; they had different numbers of syllables, different positions of the main-stressed syllable in the word, and different positions of the word in the utterance. Finally, the two words differed in whether or not, in addition to the pitch accent, they also carried the boundary-related tones (phrase accent and boundary tone) for the intonational phrase. *Dagada*, which carried no boundary tone because it occurred early in the phrase, showed a later F_0 peak alignment than *doodad*, which was the last word in the phrase and so did bear the boundary tone on its final syllable and the phrase accent before it. Because a number of studies have suggested that closely adjacent tonal targets may influence each other's alignment, in the configurational context described as tonal crowding (Silverman and Pierrehumbert, 1990; Arvaniti et al., 2006), we designed this follow-up study to determine whether tonal crowding could account for the observed differences in alignment reported. (Partial results for this follow-up study were reported earlier in (Shue et al., 2008).) The expanded corpus was carefully designed to control for factors such as vowel context, while systematically varying the number of syllables, stress pattern and structural position of the target word in the intonational phrase (early vs. late, phrase-final vs. non-final). In this way, we sought to test the hypothesis that tonal crowding from a boundary tone can result in early location of the extremum of the F_0 excursion associated with a pitch accent, and to determine whether other factors such as word length, stress pattern and early vs. late (but not final) position in the phrase have an effect. In particular, we hypothesize that the F_0 peak for a H^* accented syllable will be consistently located in or just after the accented vowel in a wide variety of contexts, i.e. for words located early or late (but not finally) in the phrase, and for words with various numbers of syllables and locations of main stress, because in these conditions there is no need to make room for boundary tone targets later in the same word. However, if the accented word is phrase-final, so that boundary tones occur in the same word, the speaker will realize the peak earlier in the accented syllable, as predicted by tonal crowding, and perhaps at a lower F_0 value as well. In addi-

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

tion, earlier results led us to hypothesize that the trough for a L* accented syllable may not show such a systematic move toward earlier alignment under conditions of crowding by immediately-following boundary tones.

In addition to effects on the alignment of the F_0 contour with the spoken words and syllables, other acoustic features of accents, such as energy levels and duration, might also be subject to tonal crowding effects. However, changes in energy measures (for example, in the presence of boundary tones) are difficult to predict. Previous studies, such as Sluijter and van Heuvan (1996a,b) and Rosenberg and Hirschberg (2006), have shown that energy measures (typically using spectral balance, intensity or banded frequency content) are correlated with the presence of stress or pitch accents, and that these measures tend to rise with the occurrence of these prosodic events. Based on these findings, we hypothesize that energy values will be higher (Rosenberg and Hirschberg, 2006) for pitch-accented syllables than for unaccented syllables; however, this difference will be smaller if a boundary tone immediately follows the accented syllable, due to falling subglottal pressure at the end of an utterance (Slifka, 2007).

We also hypothesize, based on earlier work, that pitch-accented syllables will be longer (Turk and White, 2007) than unaccented syllables in the same position in the phrase, and that word-final syllables will also be longer when they occur in phrase-final position than when they are phrase-medial (Klatt, 1976a; Beckman and Edwards, 1994; Turk and Shattuck-Hufnagel, 2007). The stimulus set is designed to test these findings from earlier work in a larger number of speakers, 10 male and 10 female. By including twenty participants in the experiment we hope to shed light on the generalizability of observations about the acoustic correlates of pitch accents in American English, both across gender and across individual speakers.

An additional pattern that we observed in the 2007 and 2008 studies is that H* accents behave somewhat differently from L*s, and we include analyses of the same type in this more extensive study. For example, we will test the hypothesis that F_0 movements are relatively less extreme for L* than for H*, that the energy increase associated with pitch accents is less for L* than for H* even when the number of pitch periods included in the measure is controlled, and that L* accents do not exhibit the same property of shifting alignment of the F_0 trough under conditions of tonal crowding as H* accents do for the peak.

2. Corpus and analysis methods

2.1. Corpus

The test corpus used in this study was carefully constructed to minimize or control for various factors, such as vowel type, syllable number and word position, which could influence the results. The corpus consists of spoken elicited utterances with specified pitch accent and boundary tone locations and types. The utterances are prosodic variations of the two sentences *Dagada gave Anne a dada* and *A dada gave Anne dagadas*, with the target words being *dagada* and *dada*. For each utterance, a single pitch accent (H* or L*) is produced on either the early target word or the late one, in either a declarative setting or an interrogative setting. The nonsense words *dagada* and *dada* were used to ensure that the lexically-stressed syllables carrying the pitch accents (*-ga-* and *da-*, respectively) had the same vowel in all cases, avoiding any vowel-specific effects. The declarative and interrogative forms of the sentences were used to elicit the phrase-final tone sequences L–L% for H* utterances and H–H% for L* utterances, on the assumption that opposite polarity for the accent vs. the pitch accent and boundary-related tones would increase the chances of obtaining clearly-detectable F_0 peaks and troughs for the accents. The eight configurations of the sentences are listed below, with the accented target word in bold font.

- ***Dagada*** *gave Anne a dada.*
  H* L–L%
- *A dada gave Anne* ***dagadas****.*
  H* L–L%
- *A* ***dada*** *gave Anne dagadas.*
  H* L–L%
- *Dagada gave Anne a* ***dada****.*
  H* L–L%
- ***Dagada*** *gave Anne a dada?*
  L* H–H%
- *A dada gave Anne* ***dagadas****?*
  L* H–H%
- *A* ***dada*** *gave Anne dagadas?*
  L* H–H%
- *Dagada gave Anne a* ***dada****?*
  L* H–H%

These same eight sentences were also recorded with the unaccented word *daily* added at the end of the sentences, to carry the boundary tone; this allowed us to determine the effects on the pitch accent realization in the late target word when the boundary tone was moved to a following word. Each of these 16 sentences was elicited with a prompt question or statement, to ensure the correct placement of the tones. For example, to elicit a H* tone on the early target word *dagada*, and a L–L% boundary tone on the unaccented late target word *dada*:

Prompt: *Was it Dagada or Daguda that gave Anne a dada?*
Response: ***Dagada*** *gave Anne a dada.*

Recordings were made for 20 native speakers of American English (10 males/10 females) between the ages of 17 and 30. For each sentence, five repetitions were recorded, for a total of 1600 utterances. The recordings were made in a sound booth at an effective sampling rate of 16 kHz. Manual segmentation of the target words *dagada* and *dada* from their context and of their main stress vowel were performed.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

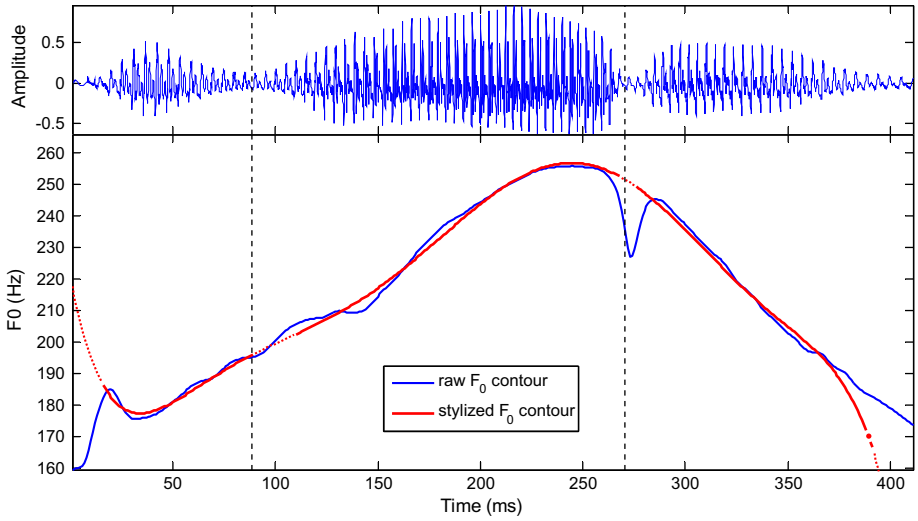

Fig. 1. Example of polynomial fitting for the target word *dagada* with a high (H*) pitch accent. The top panel shows the waveform, the bottom panel shows the raw and stylized F_0 contours. The dotted vertical lines mark the position of the manual segmentation.

2.2. Analysis methods

Measures related to the F_0 contour, energy, and duration were estimated for the analysis of the target words and main-stress syllables.

F_0 values were estimated, at a resolution of 1 ms, using the STRAIGHT algorithm (Kawahara et al., 1998) over the whole utterance. For each target word, polynomial fitting (Shue et al., 2007) was performed on the F_0 values to smooth the raw contours in order to make minima/maxima detection more accurate and robust. Similar to Kochanski et al. (2005), weighted Legendre polynomials were used for the contour approximations due to their orthogonality property. Each Legendre polynomial, $P_i(n)$, is associated with a coefficient, a_i, which enables a data vector, $y(n)$, to be approximated as $y(n) \approx \sum_{i=0}^{N} a_i P_i(n)$, where N is the desired polynomial order. For this study, we used polynomial orders from 3 to 7, where the exact order was determined by the accuracy of the resulting polynomial fit. Each word was manually checked to ensure the fitting accurately represented the raw F_0 values. Eq. (1) shows the weighted least squares error criterion, E_a, used in the optimization of the a_i's.

$$E_a = \sum \left(y(n) - \sum_{i=0}^{N} a_i P_i(n) \right)^2 \cdot W(n) \tag{1}$$

The weighted least squares error criterion was based on the signal energy, $E(n)$. In regions where the energy was relatively small, such as within the closure of /d/ and /g/, the reliability of the F_0 measures was reduced and hence, less weighting was applied. The error weighting function, $W(n)$, was determined by $E(n)$, with the threshold, E_{th}, set at a quarter of the mean word energy. After $E(n)$ drops below E_{th}, the weighting function decreases exponentially,[2] as shown in Eq. (2).

$$W(n) = \begin{cases} 1, & E(n) \geqslant E_{th} \\ e^{-(E_{th}-E(n))/E_{th}}, & E(n) < E_{th} \end{cases} \tag{2}$$

The use of the error weighting function ensures that only the most reliable F_0 values are used for the contour fitting. Examples of such fits are shown in Figs. 1 and 2 for high and low pitch-accented target words, respectively. Although raw F_0 values, in reality, are not always continuous during a voiced stop such as the /d/ and/g/ in our target syllables (i.e. microprosody), the closure duration of a voiced stop is usually small compared to the vowel duration, allowing the contour mapping to effectively smooth over these regions.

The F_0 minimum and maximum values were calculated from the smoothed contours and normalized to each speaker's mean F_0 value ($\overline{F_0}$). $\overline{F_0}$ is the average F_0 over all of that speaker's utterances. The minima and maxima were expressed as a percentage of the speaker's $\overline{F_0}$. For example, the normalized value for a particular F_0 maximum ($F_{0\max}$) is calculated as $(F_{0\max} - \overline{F_0})/\overline{F_0} \times 100\%$.

Energy measures were calculated using an adaptive window size to account for the effects of F_0. The window size at a particular point in time was set to three pitch periods, as determined by the STRAIGHT estimated F_0 value at that point. For example, the F_0 value at the H* peak in Fig. 1 has a value of 255 Hz; with a sampling frequency of 16 kHz, this leads to a window size of $\lfloor 16{,}000/255 \times 3 \rfloor = 188$ samples. Utterances were normalized to have the same maximum energy value. The energy of each syllable is normalized with respect to the utterance's mean energy value and then used in ANOVA tests.

The duration of each main-stressed vowel was obtained from the manual segmentation. Onset and offset times were taken at the points where there was evidence of syllable closure or release, such as the abrupt fall in signal amplitude or the sudden loss of signal periodicity. An example of two such points is shown in Fig. 1 as two vertical dotted lines.

[2] Other functions were tried, such as linear and piece-wise linear functions, but the exponential function provided the best performance due to its smooth roll-off.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

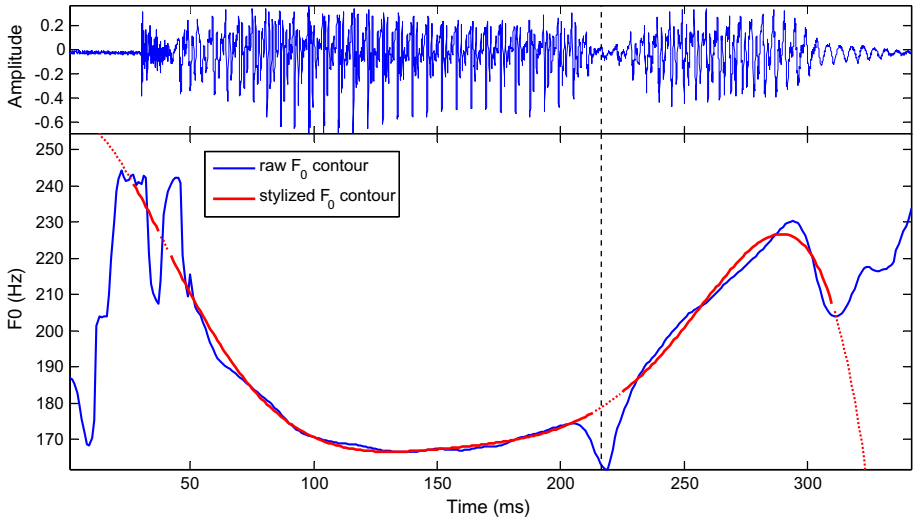

Fig. 2. Example of polynomial fitting for the target word *dada* with a low (L*) pitch accent. The top panel shows the waveform, the bottom panel shows the raw and stylized F_0 contours. The dotted vertical line marks the position of the manual segmentation.

The results of the analyses were grouped according to the gender of the speakers. This was done due to the well-known physiological and acoustical differences between male and female speakers (Titze, 1989). Furthermore, it was shown in Iseli et al. (2007) that many measures related to the voice source were dependent on the value of F_0 and thus, may be attributed to the gender differences.

3. Results

We focused on the vowels in the main-stress syllables of the target words, which had relatively clear boundaries. We distinguish between several properties of the analyzed syllables: early vs. late position of the target word in the utterance; for late position, boundary vs. non-boundary position; position of the lexically-stressed syllable in the target word (medial in *dagada*, initial in *dada*); accentedness (accented vs. non-accented) and if accented, whether the accent was H* or L*. No vowel effects were examined because the same vowel /a/ was intentionally used in all target syllables.

Four types of positions were examined, illustrated here with the labels *no-bnd-early*, *bnd*, *no-bnd-early-daily* and *no-bnd-late-daily* for declarative sentences with the target word, *da**ga**da*, where the –***ga***– syllable is always stressed and can be accented or not:

(1) *no-bnd-early*: *Da**ga**da gave Anne a dada.*
(2) *bnd*: *A dada gave Anne da**ga**das.*
(3) *no-bnd-early-daily*: *Da**ga**da gave Anne a dada daily.*
(4) *no-bnd-late-daily*: *A dada gave Anne da**ga**das daily.*

Additionally, in this section, for notation purposes, we will refer to all non-boundary (*no-bnd-early*, *no-bnd-early-daily* and *no-bnd-late-daily*) cases as *no-bnd* and all unaccented cases as *no-acc*.

Analysis of variance (ANOVA) tests were performed using the software package SPSS (v. 16.0) to check for the statistical significance of the results. The two fixed factors, speaker and tone (H*/L*), or speaker and boundary (yes/no, and if yes, H–H%/L–L%) were used to examine the effects on the measures. Tests where the null hypothesis has a probability of $p < 0.001$ were considered to be statistically significant.

We present the results based on our hypotheses made earlier, that is (1) tonal crowding affects the position and height of the F_0 maxima/minima of a pitch-accented syllable, and (2) phrase-final lengthening is increased when the phrase-final word also includes an accent. We also

Table 1
Position of the F_0 peak/trough as a percentage of the speaker's vowel duration. The results shown are averaged for the male and female speakers for the target words *dagada* and *dada*; standard deviation values are shown in parentheses. The statistical significance (s.s.) column shows the ANOVA results for *no-bnd* vs. *bnd*. For significant results, the F (ratio of the model mean square to the error mean square) and η^2 (measure of the effect size) values are given.

F_0 peak/trough position mean (std.) in %		Males			Females		
		no-bnd	*bnd*	s.s. $F(1,180)/\eta^2$	*no-bnd*	*bnd*	s.s. $F(1,180)/\eta^2$
dagada	H*	85(17)	65(13)	**77.1/.300**	92(15)	69(13)	**222.8/.537**
	L*	58(14)	52(14)	No	59(13)	58(19)	No
dada	H*	83(13)	70(17)	**40.0/.188**	90(15)	68(13)	**139.6/.429**
	L*	51(14)	48(19)	No	55(11)	54(14)	No

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

Table 2
Height of the F_0 excursion as a percentage of the speaker's mean F_0. Average results are shown for the *no-bnd* vs. *bnd* conditions for the male and female speakers for the target words *dagada* and *dada*; standard deviation values are shown in parentheses.

ΔF_0 mean (std.) in %		Males			Females		
		no-bnd	*bnd*	s.s. $F(1,180)/\eta^2$	*no-bnd*	*bnd*	s.s. $F(1,180)/\eta^2$
dagada	H*	24(20)	8(17)	**78.7/.305**	35(27)	23(28)	**46.3/.194**
	L*	−26(12)	−29(12)	No	−35(9)	−40(16)	No
dada	H*	21(19)	8(17)	**41.2/.193**	36(25)	23(28)	**41.1/.181**
	L*	−27(12)	−29(11)	No	−35(8)	−42(11)	**28.6/.137**

examined energy measures to determine the acoustic effects of pitch accents and boundary tones on this parameter.

3.1. F_0

To test the hypothesis that tonal crowding has an effect on the acoustic measures of pitch-accented syllables, we first analyzed the F_0 contours of pitch-accented syllables in words which contained a boundary tone (*bnd*) vs. those not containing a boundary tone (*no-bnd*). To separate out the effects of the serial position of the word, we then analyzed the F_0 measures for the *bnd* vs. *no-bnd-early*(*-daily*) cases and the *bnd* vs. *no-bnd-late-daily* cases.

Tables 1 and 2 show the results of ANOVA tests when F_0 peak/trough positions and heights are tested against the fixed factors *no-bnd* and *bnd* for the target word *dagada* and *dada* respectively. For statistically significant (s.s.) results, the F ratio[3] and partial η^2 (measure of effect size) values are also given.

In Table 1, the values in the rows represent the mean F_0 peak/trough positions relative to the duration of the target vowel, where 50% corresponds to the middle of the target vowel. For example, for male speakers, the H* accented *dagada* for the cases where the target word did not also include the boundary tone (i.e. the *no-bnd* case) had an F_0 peak which occurred, on average, at a position which was 85% of the mean vowel duration with a standard deviation of 17%. This is in contrast to the case where the target word also carried the boundary tone (*bnd*), which had a position of 65% of the vowel duration. No statistically significant effects were observed for L* accented target words. The results are similar for the target word *dada*, again showing the shift of the peak to an earlier position in the presence of the boundary tone.

In Table 2, the values in the rows denote the mean F_0 peak/trough excursion relative to the mean of the speaker's F_0 values in percent; i.e. 0% corresponds to a peak/trough exactly at the speaker's mean F_0 calculated over all the speaker's utterances. For example, for the female speakers producing *dada* in the non-boundary condition (*no-bnd*), the H* accent resulted in an F_0 peak which was on average 36% higher than the speaker's mean F_0, while for the boundary case (*bnd*), the F_0 peak was only 23% higher, showing a lesser excursion of F_0 for the pitch accent in the presence of a boundary tone on the same word. Note that the boundary tone for a H*/L* accented target word is L–L% and H–H% respectively.

The ANOVA tests show that regardless of gender, for the H* accented vowels the *no-bnd*/*bnd* factors have a statistically significant effect on both the position and height of the F_0 peak. For both target words, the *no-bnd* case had an F_0 peak position which occurred much later than the *bnd* case and peak heights that were greater. Interestingly, the L* accented vowels showed no statistically significant shift in height or duration, with the exception of *dada* for female speakers for the *height* factor, where a weak effect size ($\eta^2 = 0.137$) is observed.

Since the *no-bnd* case contains instances where the pitch-accented word is at the start of the sentence (*no-bnd-early* and *no-bnd-early-daily*) and near the end of the sentence (*no-bnd-late-daily*), it is possible that the serial position of the word may also affect the results, although this was not hypothesized. For example, if F_0 declination occurs over the course of the utterance, the F_0 peak for a H* may be lower for an accent that occurs late in the utterance

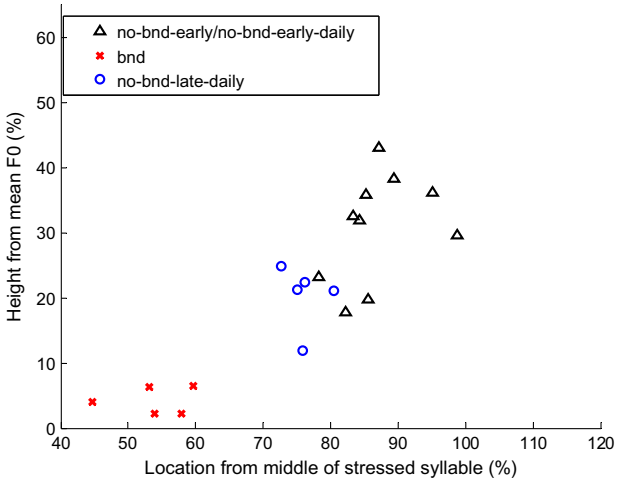

Fig. 3. Scatter plot for the target word *dagada* showing relative F_0 peaks for H* and their relative positions in the accented target vowel for a male speaker in three different contexts: 1. *no-bnd-early*/*no-bnd-early-daily* (triangles); 2. bnd (crosses); 3. no-bnd-late-daily (circles).

[3] The F value is defined as the ratio of the model mean square to the error mean square and the partial η^2 value is calculated as $SS_{effect}/(SS_{effect} + SS_{error})$, where SS_{effect} is the sum of squares of the effect and SS_{error} is the sum of squares of the error.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

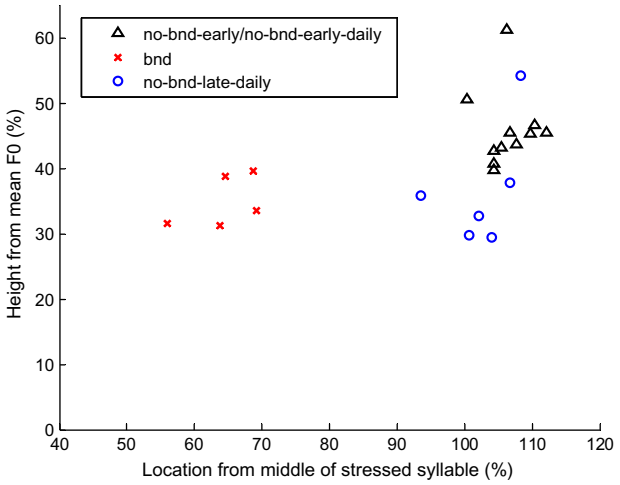

Fig. 4. Scatter plot for the target word *dada* showing relative F_0 peaks for H* and their relative positions in the accented target vowel for a female speaker in the three different contexts.

than for an accent that occurs early. To confirm that tonal crowding rather than declination is the main cause of the lower F_0 peak, we also analyzed the F_0 contour peak and height by comparing *bnd* vs. *no-bnd-early*(*-daily*) and *bnd* vs. *no-bnd-late-daily*. Figs. 3 and 4 show the F_0 peak positions for H* for a typical male/female speaker producing *dagada*/*dada* respectively in the three different contexts; note that the *no-bnd-early* and *no-bnd-early-daily* cases showed similar results, hence were considered together. For these two speakers, it can be seen that the *bnd* cases generally had earlier F_0 peaks compared to both of the other two cases, indicating the temporal crowding effect of the boundary tone. The male speaker (Fig. 3) also had F_0 peaks that were significantly lower for the *bnd* case than the other two cases, while for the female speaker (Fig. 4) the F_0 peaks for the *bnd* case was lower than the *no-bnd-early*(*-daily*) case and similar to the *no-bnd-late-daily* case, indicating the possible compression effects of the boundary tone. These trends were observed for 8/10 male speakers and 9/10 female speakers. Note that the ANOVA tests in this section were carried out using the results from all speakers, including those that did not conform to the general trend.

Statistical analysis comparing F_0 peak/trough position and height for the *bnd* vs. *no-bnd-early*(*-daily*) conditions are shown in Tables 3 and 4 respectively, for target words *dagada*/*dada*. Similar to the trends shown in Tables 1 and 2 for the *bnd*/*no-bnd* comparison, it can be seen that for the H* pitch accent, the position and height of the F_0 peak are affected significantly by the presence/absence of the boundary tone, with the *no-bnd-early*(*-daily*) peak occurring later and higher than the *bnd* case. For example, for female speakers the H* accented target word *dagada* showed a peak position difference of 27% (from 69% to 96%) and a height difference of 18% (from 23% to 41%) when comparing the *bnd* case to the *no-bnd-early*(*-daily*) cases. L* accented words did not exhibit any statistical significance for the *position* measure, except for male speakers for the target word *dagada* which showed a small effect size ($\eta^2 = 0.104$). Similarly, although the *height* measure differences for L* accented target words are statistically significant, the difference between the means for the *bnd* and *no-bnd-early*(*-daily*) conditions are small with a relatively weak effect size.

Tables 5 and 6 show the F_0 results when the target words are tested for the *no-bnd-late-daily* vs. *bnd* effect for the position and height measures, respectively. Similar to the previous results, it can be seen that, regardless of gender and target word, the H* accented F_0 peak, on average, occurred much later for the *no-bnd-late-daily* case and these results were statistically significant. Height differences were less consistent than position differences, with only the male speakers showing statistical significance for the target word *dagada*. For L* accented syllables, only female speakers for the height measure on the target word *dada* showed any statistical significance, and the effect size is relatively weak.

These results are consistent with our hypothesis that the F_0 peak for an H* accent is usually located towards the end or just after the accented vowel, unless the need to realize other tonal targets (such as a boundary tone) on the same word causes the speaker to realize the peak earlier in the accented syllable, presumably in order to make room for the realization of the additional targets. This further supports the hypothesis that the location of the F_0 peak for an H* is affected more by tonal crowding (Arvaniti et al., 2006) from boundary tones than by the mere serial position of the accented word, or by its number of syllables. Interestingly, the F_0 peak position on average is later for the female speakers than for the males; this may be due to larger relative change in F_0 which could require more time to achieve. The lack of a consistent trend for L* pitch-accented syllables is as predicted, and suggests the implementation of L*

Table 3
Position of the F_0 peak/trough as a percentage of the speaker's target vowel duration. Results shown are average values for the male and female speakers for target words *dagada* and *dada* in the *no-bnd-early*(*-daily*) vs. *bnd* condition; standard deviation values are shown in parentheses.

F_0 peak/trough position mean (std.) in %		Males			Females		
		no-bnd-early(-daily)	*bnd*	s.s. $F(1,270)/\eta^2$	*no-bnd-early*(-daily)	*bnd*	s.s. $F(1,270)/\eta^2$
dagada	H*	88(17)	65(13)	**167.3/.565**	96(15)	69(13)	**292.1/.678**
	L*	59(14)	52(14)	**14.9/.104**	61(13)	58(19)	No
dada	H*	84(11)	70(17)	**36.0/.229**	93(15)	68(13)	**171.1/.563**
	L*	52(13)	48(19)	No	55(9)	54(14)	No

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

Table 4
Height of the F_0 excursion as a percentage of the speaker's mean F_0. Results shown are for the *no-bnd-early*(*-daily*) vs. *bnd* conditions; standard deviation values are shown in parentheses.

ΔF_0 mean(std.) in %		Males			Females		
		no-bnd-early(*-daily*)	*bnd*	s.s. $F(1,270)/\eta^2$	*no-bnd-early*(*-daily*)	*bnd*	s.s. $F(1,270)/\eta^2$
dagada	H*	26(19)	8(17)	**71.6/.372**	41(22)	23(28)	**113.8/.461**
	L*	−24(11)	−29(12)	**32.8/.204**	−34(7)	−40(16)	**19.3/.130**
dada	H*	26(19)	8(17)	**71.6/.372**	41(22)	23(28)	**113.8/.461**
	L*	−26(12)	−29(11)	**45.8/.259**	−35(8)	−42(11)	**23.6/.153**

Table 5
Relative position of the F_0 peak/trough as a percentage of the speaker's target vowel duration. Results shown are average vales for the male and female speakers for target words *dagada* and *dada* in the *no-bnd-late-daily* vs. *bnd* condition; standard deviation values are shown in parentheses.

F_0 peak/trough position mean (std.) in %		Males			Females		
		no-bnd-late-daily	*bnd*	s.s. $F(1,183)/\eta^2$	*no-bnd-late-daily*	*bnd*	s.s. $F(1,183)/\eta^2$
dagada	H*	80(17)	65(13)	**55.5/.404**	86(13)	69(13)	**202.2/.697**
	L*	55(13)	53(14)	No	56(13)	58(19)	No
dada	H*	80(15)	70(17)	**11.6/.125**	83(14)	68(13)	**117.6/.586**
	L*	48(14)	48(19)	No	54(13)	54(14)	No

Table 6
Relative height of the F_0 excursion as a percentage of the speaker's mean F_0. Results shown are for *no-bnd-late-daily* vs. *bnd*; standard deviation values are shown in parentheses.

ΔF_0 mean(std.) in %		Males			Females		
		no-bnd-late-daily	*bnd*	s.s. $F(1,183)/\eta^2$	*no-bnd-late-daily*	*bnd*	s.s. $F(1,183)/\eta^2$
dagada	H*	19(23)	8(17)	**49.4/.376**	24(29)	23(28)	No
	L*	−29(11)	−29(12)	No	−38(11)	−40(16)	No
dada	H*	13(16)	8(17)	No	26(27)	23(28)	No
	L*	−30(11)	−29(11)	No	−36(8)	−42(11)	**15.2/.159**

accents may be governed by different principles from those governing H* accents. These results also highlight the fact that cues which are found to be correlated with H* pitch accents may not necessary be correlated with L* accents; so that it is important to separate pitch accents into H* and L* categories when analyzing their correlates.

The results for the height measure for the H* accented syllables were seen to be more consistent for utterances which had the accented syllable at the start of the sentence (i.e. *no-bnd-early(-daily)*). For these cases, the F_0 peak was consistently higher than for the boundary cases. Although this trend was also seen for the utterances which had the late accented syllable (*no-bnd-late-daily*), those results were not statistically significant. This may be an effect of the need to realize the final low boundary tone (L–L%) on the word *daily*, which immediately follows the pitch-accented word. Another possible explanation is that the height of the F_0 peak for the late accented syllable may be influenced by the natural F_0 declination which can occur over the course of declarative statements.

3.2. Energy

We hypothesized that the energy change for a pitch accent might be similar across gender and pitch accent type, but different in boundary vs. non-boundary conditions, because of respiratory and subglottal pressure changes at the end of an utterance (Slifka, 2007). Figs. 5 and 6 show, respectively, the scatter plots for male/female speakers of the relative mean energies for the H* accented vowel where the target word is *dagada*/*dada* in the three conditions, *no-bnd-early(-daily)*, *bnd*, and *no-bnd-late-daily*. The *y*-axis represents the change in relative energy in relation to the average energy of each utterance in percent, so a 100% value would correspond to twice the average utterance energy. It can be seen that for most speakers the boundary (*bnd*) case provides, on average, the least change in energy compared with the non-boundary cases (*no-bnd-early, no-bnd-early-daily* and *no-bnd-late-daily*). The considerable overlap between the *no-bnd-daily* and the *bnd* cases suggest that, for some speakers, the fall in energy could occur as early as the accented syllable of a penultimate word.

Two-way ANOVA results for the energy measure as a function of gender, pitch accent tone-type and presence of adjacent boundary tones are shown in Tables 7 and 8 for target word *dagada* and *dada*, respectively. The values represent the percentage of change of the mean target syllable energy from the mean utterance energy. For example, the first row in Table 7 shows that the average energy of the

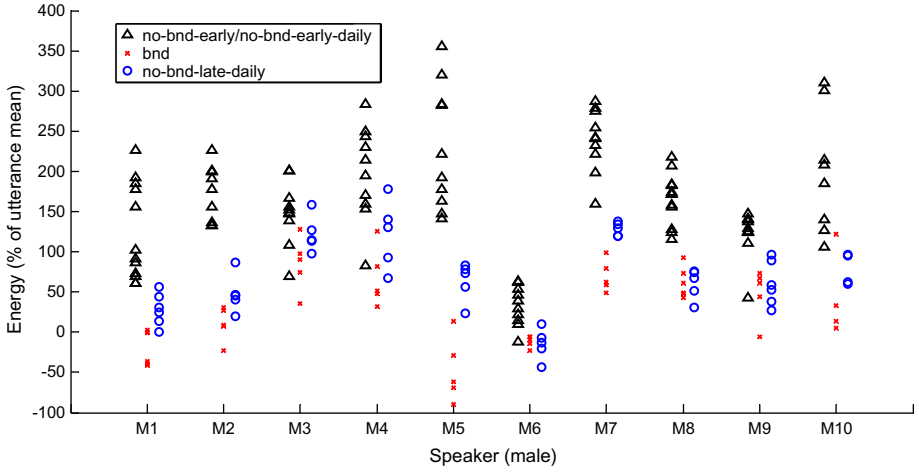

Fig. 5. Scatter plot of relative mean vowel energy of the H* accented vowel for the target word *dagada* for all male speakers in three different contexts: 1. *no-bnd-early*/*no-bnd-early-daily* (triangles); 2. *bnd* (crosses); 3. *no-bnd-late-daily* (circles).

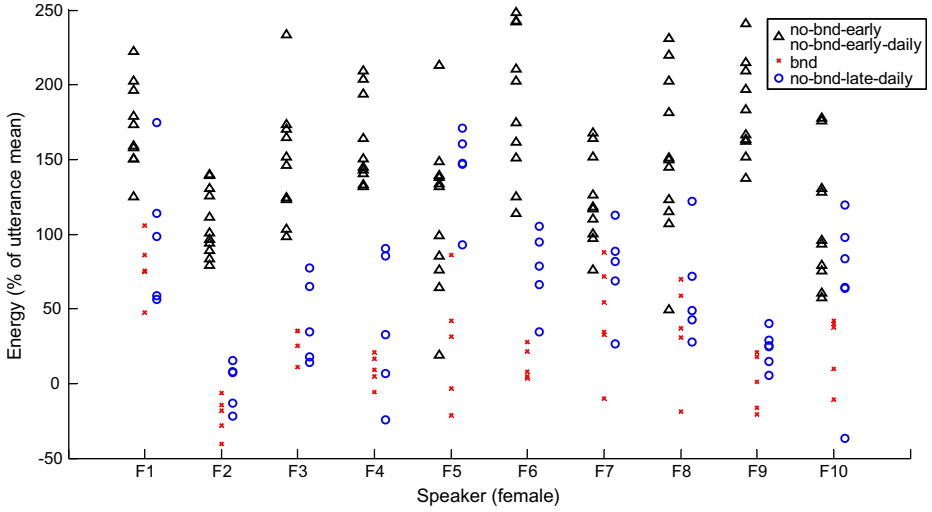

Fig. 6. Scatterplot of relative mean vowel energy of the H* accented vowel for the target word *dada* for all female speakers in three different contexts: 1. *no-bnd-early*/*no-bnd-early-daily* (triangles); 2. *bnd* (crosses); 3. *no-bnd-late-daily* (circles).

Table 7
Relative energy mean, standard deviation (std.) in parenthesis, of stressed syllables for the target word *dagada* for male and female speakers. Results are shown for *no-bnd* vs. *bnd*.

Energy mean(std)	Males			Females		
	no-bnd	*bnd*	s.s. $F(1,185)/\eta^2$	*no-bnd*	*bnd*	s.s. $F(1,185)/\eta^2$
H*	131(80)	30(51)	**114.2/.382**	125(74)	9(39)	**132.9/.409**
L*	36(72)	−10(49)	**39.1/.172**	−11(57)	−54(34)	**31.4/.147**
no-acc (H–H%)	36(52)	−17(38)	**50.9/.215**	35(48)	13(40)	**14.2/.073**
no-acc (L–L%)	3(61)	−67(26)	**68.6/.271**	4(65)	−79(13)	**94.9/.338**

H* accented vowel for male speakers in the *no-bnd*/*bnd* case is 131/30% higher than the mean energy of the utterance, statistically significant with an *F*-value of 114.2 and effect size of 0.382. It can be seen that on average, the change in energy was lower for the *bnd* case than for the *no-bnd* case regardless of gender and pitch accent. This trend also extends to the non-accented cases (*no-acc*) for both the interrogative (H–H%) and declarative (L–L%) utterances. Interestingly, L* accented syllables had a lower energy change than H* accented syllables and for female speakers, the energy change, on average, was significantly lower than even the *no-acc* cases. There were also significant differences associated with pitch accent type; for example, energy values generally showed greater variance for males

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

Table 8
Relative energy mean, standard deviation (std.) in parenthesis, of stressed syllables for the target word *dada* for male and female speakers. Results are shown for *no-bnd* vs. *bnd*.

Energy mean(std)	Males			Females		
	no-bnd	*bnd*	s.s. $F(1,185)/\eta^2$	*no-bnd*	*bnd*	s.s. $F(1,185)/\eta^2$
H*	140(68)	38(42)	**178.0/.490**	116(63)	24(34)	**108.7/.369**
L*	34(61)	−12(45)	**49.4/.210**	−30(49)	−50(32)	**16.0/.081**
no-acc (H–H%)	32(58)	−3(50)	**20.6/.099**	31(46)	14(36)	No
no-acc (L–L%)	5(60)	−66(32)	**71.1/.278**	5(65)	−75(17)	**91.0/.322**

across boundary conditions, and values were higher for H*/*no-acc*(H–H%) than for L*/*no-acc*(L–L%) regardless of gender. Thus, as with the F_0 peak location, the results for energy show the importance of context information in understanding the acoustic correlates of tonal targets.

3.3. Duration – effects of pitch accent on phrase-final lengthening

Predictions about the effects of pitch accent on phrase-final lengthening are complex, since several different factors are at work. These include duration lengthening associated with main lexical stress (Beckman and Edwards, 1994), with the main-stress syllable of the phrase-final word (Turk and Shattuck-Hufnagel, 2007), with the final syllable of the phrase (Klatt, 1976b) and with a pitch accent (Turk and White, 2007). We hypothesized that phrase-final lengthening might increase with the addition of a pitch accent on the phrase-final word, possibly in order to allow the speaker more time to achieve both prosodic targets. In this section, we first test the effects of high and low pitch accents on duration and target word position. We then test the phrase-final-lengthening effects in our corpus by comparing the final syllable durations for the non-accented words *dagada* and *dada* in the *late* vs. *early* vs. *bnd* conditions; the *late* case is where the unaccented word is positioned before the word *daily*, the *early* case is where the unaccented word appears at the start of the utterance and the *bnd* case is where the unaccented word is the phrase-final word. The main-stress-syllable lengthening effect of the phrase-final word is then tested by comparing the stressed-syllable durations of the non-accented *dagada* and *dada* in the *late* vs. *early* vs. *bnd* condition. Finally, the hypothesized extra phrase-final lengthening effect is checked by comparing the final-syllable durations of the target words for the *no-bnd-early* vs. *bnd* conditions.

Fig. 7 shows the average durations, for male and female speakers, of the main-stressed syllable (-ga-) for the target word *dagada* in the early, late and boundary conditions. It can be seen that regardless of word position, the average duration for the Non case case is much lower than for either L* or H* cases, confirming the durational lengthening associated with pitch accents (Turk and White, 2007). While ANOVA tests on the tone-type showed that the results are statistically significant, post-hoc analyses revealed that the L* and H* accented durations are virtually indistinguishable. This result means that while durational lengthening occurs in the presence of pitch accents, it is not affected by the pitch-accent type. Furthermore, the results also show that the boundary position has on average longer durations for the Non, L* and H* cases than for both the early and late positions. Results for the target word *dada* are similar.

Fig. 8 shows the comparisons of the average durations and the corresponding error bars of the final syllable (*daga**da*** and *da**da***) for male and female speakers. The effects of phrase final lengthening can be clearly seen in the longer average duration for the *boundary* case. This

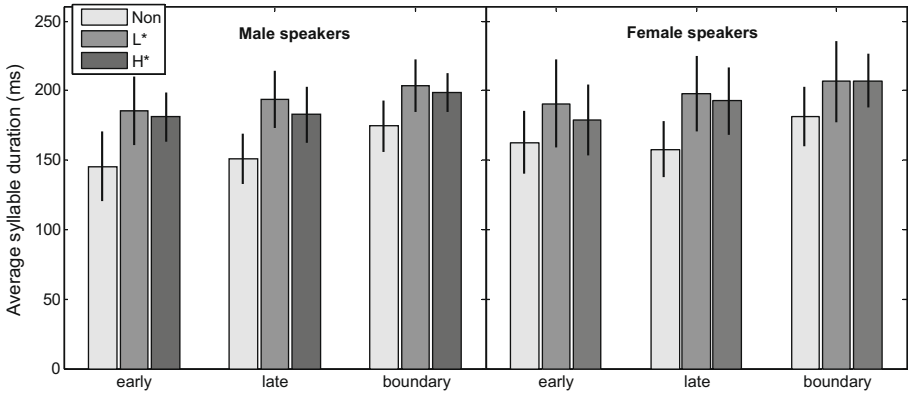

Fig. 7. Average main-stressed-syllable duration with no (Non), L*, and H* pitch accents for male and female speakers for the target word *dagada* in the early, late and boundary positions.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

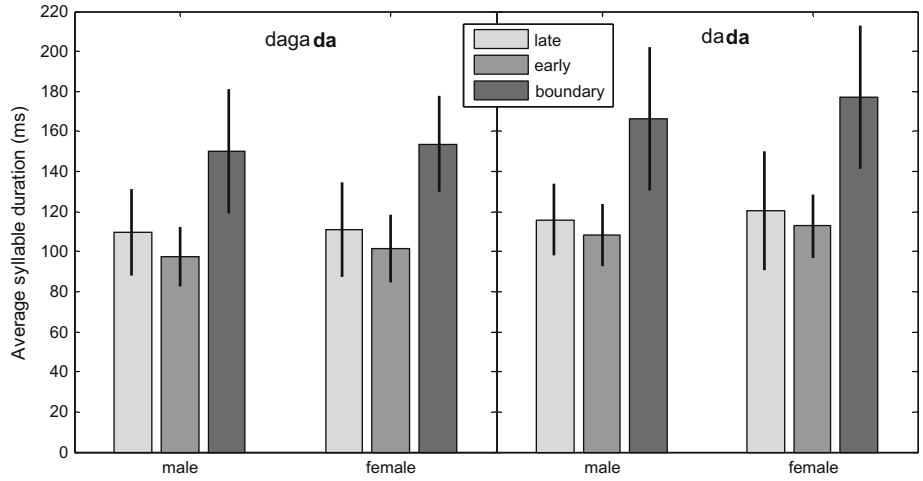

Fig. 8. Average final vowel durations and error bars of the unaccented words *dagada* and *dada* for male and female speakers in the *late*, *early* and *boundary* positions. The increased duration for the *boundary* case confirms the effects of phrase final lengthening.

result is statically significant for both words when the fixed factors, speaker and word position, are used in an ANOVA test; for male speakers, $F(2/2, 381/383) = 413/414$, $p = 0.000/0.000$, and $\eta^2 = 0.68/0.68$ and for female speakers, $F(2/2, 377/384) = 489/486$, $p = 0.000/0.000$ and $\eta^2 = 0.72/0.72$ for the unaccented words *dagada*/*dada*. Note that the two no-boundary cases, *early* and *late*, are not significantly different in their averaged final-syllable durations, indicating that for the *late* case, the final vowel may not be close enough to the boundary for boundary-related effects to appear. This may be because it is too far away in time or because of the intervening word boundary.

Fig. 9 shows the comparisons of the average durations and the corresponding error bars of the main-stressed syllable (*da**ga**da* and ***da**da*) for male and female speakers, when the target words are unaccented and in the *late*, *early* and *boundary* conditions. It can be seen that the boundary case has the largest average duration for both target words and for both genders. These results were all statistically significant ($p < 0.001$) and confirm the lengthening effect of boundaries on unaccented main-stress syllables reported in Turk and Shattuck-Hufnagel (2007).

Figs. 10 and 11 show a comparison of the average durations of the final vowel for male and female speakers for the target word *dagada* and *dada* respectively, for three conditions: no accent on the preceding syllable, H^* accent on the preceding syllable, and L^* accent on the preceding syllable. On average, the duration of the final vowel increased when there was a preceding pitch accent. However, when the pitch accent was further categorized into H^* and L^* cases, it can be seen that there are some speakers who differ from the general trend; for example, speaker M1 and speaker F5 showed a shorter duration for the H^*-preceding condition than for the no-accent condition. Statistical significance tests of the phrase-final vowel durations against the presence/absence of accent (and if present,

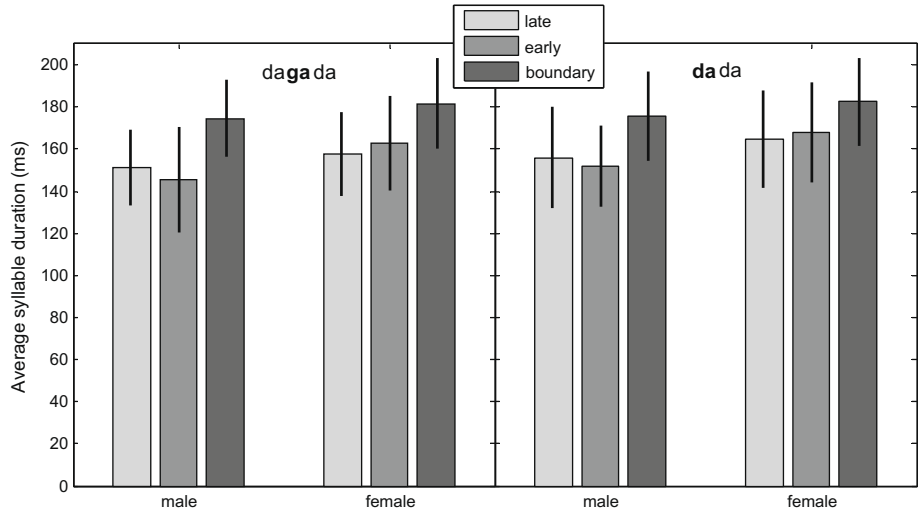

Fig. 9. Average main-stressed vowel durations and error bars of the unaccented words *dagada* and *dada* for male and female speakers in the *late*, *early* and *boundary* positions. The increased duration for the *boundary* case confirms the unaccented main-stressed syllable lengthening at the boundary condition.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

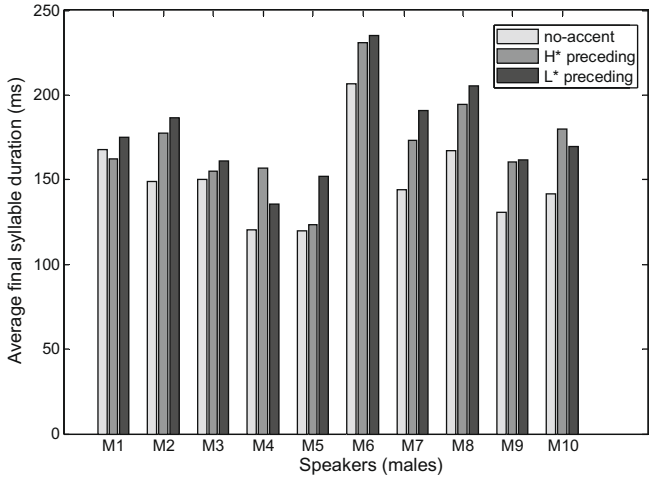

Fig. 10. Average final-syllable durations for male speakers for phrase final target word *dagada* with no preceding accents, with a preceding H* accent, and with a preceding L* accent.

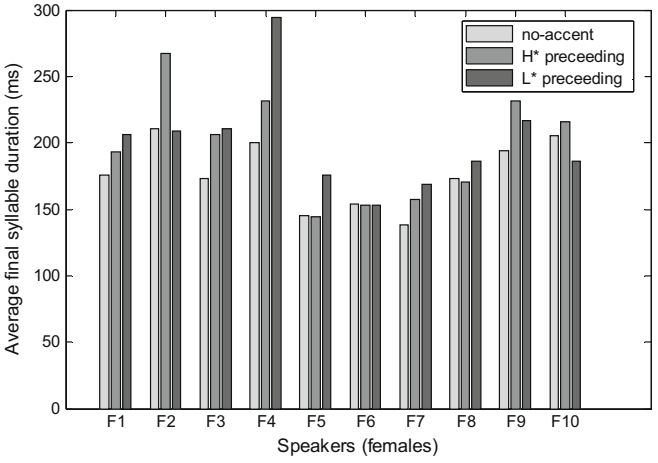

Fig. 11. Average final-syllable durations for female speakers for phrase final target word *dada* with no preceding accents, with a preceding H* accent, and with a preceding L* accent.

H*/L*) confirm the general trends seen in the figures; the ANOVA results are shown in Table 9 and the results were statistically significant for both genders and target words. Interestingly, the mean durations show that on average, the L*-preceding case was marginally longer than the H*-preceding case and much longer than the no-accent case. This result indicates that while there is a slight difference between the L*-preceding and H*-preceding cases, the extra phrase-final lengthening is more dependent on the presence/absence of a preceding accent than on the type of accent.

4. Discussion

Our goals in this paper were to test the hypothesis that tonal crowding contributed to the striking difference in the alignment of the F_0 peak associated with H* in phrase-final vs non-final words in Shue et al. (2007, 2008), and to explore other aspects of the acoustic correlates of American English H* and L* pitch accents. In this section we first review the implications of our results for these goals (4.1), then discuss the additional insights that emerge from access to results for 20 individual speakers (4.2), and finally discuss in more general terms the concept of tonal crowding (4.3): how this term has been used, the types of alignment effects it has been invoked to account for, and potential future steps toward a more comprehensive theory of how adjacent tones influence each other.

4.1. Overall results

In this section we briefly review the significance of our findings with respect to the acoustic correlates of F_0, energy and duration. Our analysis of F_0 correlates tested the tonal crowding hypothesis, which predicts that the F_0 peak for a H* pitch accent will be located in approximately the same place with respect to the accented syllable in all of our conditions, except the one in which the H* occurs on the final syllable of the phrase; in this case the peak is predicted to occur earlier, because of the crowding effect from the phrase accent and boundary tone which must be realized later in the same word. This prediction was supported by the results: the peak location was not significantly different across the two target words with different numbers of syllables (*dagada* and *dada*), across the two non-final locations (early and late), across genders and for 17/20 individual speakers. In contrast, the peak was located significantly earlier in the accented syllable when the accent occurred on the phrase-final word (i,e. in the boundary condition),

Table 9
Average duration, standard deviation (in parenthesis) of the final syllable of *dagada*/*dada* with no preceding pitch accent, with a H* preceding accent, and with a L* preceding accent. All results were statistically significant.

Mean duration (ms)	Males			
	No accent	H* preceding	L* preceding	s.s. $F(2,174)/\eta^2$
dagada	150 (31)	172 (30)	177 (32)	**49.1/.361**
dada	166 (36)	179 (38)	208 (47)	**57.8/.398**
Mean duration (ms)	Females			
	No accent	H* preceding	L* preceding	s.s. $F(2,175)/\eta^2$
dagada	154 (24)	175 (35)	185 (34)	**81.6/.480**
dada	177 (36)	196 (50)	201 (40)	**16.8/.163**

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

as predicted by the crowding hypothesis. Thus the results for H* support our hypothesis that in these utterances tonal crowding on the target word tends to shift the H* F_0 peak to occur earlier, presumably to allow room for the boundary tone to be realized. The height of the F_0 peak for H* accents was also reduced for target words that occurred late in the phrase (i.e. in the *bnd* and *late-no-bnd* conditions, compared to the early condition). This is to be expected if these utterances were produced with overall global declination in F_0. An additional lowering of the F_0 peak was found for the *bnd* condition over the *late-no-bnd* condition, suggesting a possible truncation of the F_0 rise by the following L–L% tone combination. Note that truncation can apply to both the time domain and frequency domain.

Interestingly, the F_0 troughs associated with L* accents did not follow this pattern of proportionally earlier location in the syllable in response to tonal crowding; no significant differences were found between the *bnd* and the two *no-bnd* conditions. This raises the possibility that L*s are more variably realized, or perhaps are governed by a different set of principles than H* accents. We note that Arvaniti and Garding (2007) found that the F_0 trough for the L in their L + H* accents was aligned more consistently than the F_0 peak for the H*, which they also view as evidence that L and H tones have different properties. Similarly, Arvaniti et al. (2006) found significant differences in the behavior of the H and L targets in their study of Greek Polar Questions (see Section 4.3 below for further discussion).

Our predictions for comparisons of energy levels in accented syllables were that, as in earlier work, accented syllables would have higher energy levels, but that the difference might be less in the *bnd* condition because of falling subglottal pressure in the phrase-final word. The results were consistent with this prediction: energy differences between accented and unaccented syllables were smaller on average for the *bnd* case than for the other cases, regardless of gender and pitch accent. This is consistent with trends found in Shue et al. (2008). The energy difference between accented and unaccented syllables is less on average for L* than for H*, and less for L–L% compared to H–H%. This result is somewhat surprising because our analysis method, which takes a window of three pitch periods, was designed to neutralize any effects of F_0 levels themselves by using pitch-synchronous energy. Thus the smaller energy difference for L* and for L–L%, like the results for F_0 alignment, raises the possibility that L targets are governed by different principles than those that govern H targets.

Our results for duration showed that phrase-final-syllable durations were, on average, longer when the immediately preceding syllable carried a pitch accent. This may be because the duration increase associated with a pitch accent can extend into the following syllable, as reported by Turk and White (2007). Furthermore, it was found that, on average, syllables with preceding L* accents have longer duration than syllables with preceding H* accents. The longer duration attributed to syllables with preceding L* accents may be due to the shorter time required to achieve the preceding falling pitch. This durational difference in high and low pitch changes has been reported in Ohala and Ewan (1973); Sundberg (1979) and was later confirmed in Xu and Sun (2002). A shorter time required to complete the preceding L* accent leaves more time for the final syllable. Similarly the longer time required to achieve a rising pitch can be attributed to the subsequent shorter duration of a final syllable which has a preceding H* accent. This result is yet one more indication that L* and H* pitch accents are not implemented in the same way. Speaker-specific differences were also found, with some speakers not conforming to the general trends of longer duration for L* vs. H* and shortest duration for the case without a preceding pitch-accented syllable.

Our duration results are in line with Beckman and Edwards (1994), Klatt (1976b), Turk and Shattuck-Hufnagel (2007) and Turk and White (2007), among others, who found lengthening of the phrase-final syllable. In addition to lengthening of the phrase-final syllable, we found boundary-related lengthening in the main-stress (penultimate) syllable of both *dada* and *dagada* even when these words were unaccented. This finding extends the results for main-stress-syllable lengthening in phrase-final position utterance-medially (Turk and Shattuck-Hufnagel, 2007) to main-stress-syllable lengthening in utterance-final position.

Overall, our results provide evidence that the acoustic correlates of high and low pitch accents in American English include F_0, energy and duration. By analyzing results separately for H* and L*, we have added to the growing evidence that these two kinds of pitch accents do not behave in precisely the same ways. In particular, the lack of evidence for effects of tonal crowding for L*s highlights the fact that the parameters of such crowding have not been thoroughly explored. We discuss some of the requirements for a full theory of tonal interaction in the final section of the paper. Before turning to that discussion, however, we examine the significance of the varying results for individual speakers.

4.2. Individual speaker analyses

The results reported above have the advantage of being normalized for individual speakers, potentially increasing the power of the analyses. In addition, the availability of 20 sets of individual results gives some estimate of the range of variation in the acoustic correlates of tonal targets across speakers, and of the differences in response to contextual factors such as tonal crowding. These differences are as important as the averaged trends, because they show the difficulty of placing some speakers into generalized models. The F_0 results, shown in Section 3.1, confirmed the hypothesis that the F_0 peak of an H* accented syllable would shift to an earlier point if there was another prosodic target (in this case, the boundary-related tones) which needed to be realized on the same word. This trend was

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

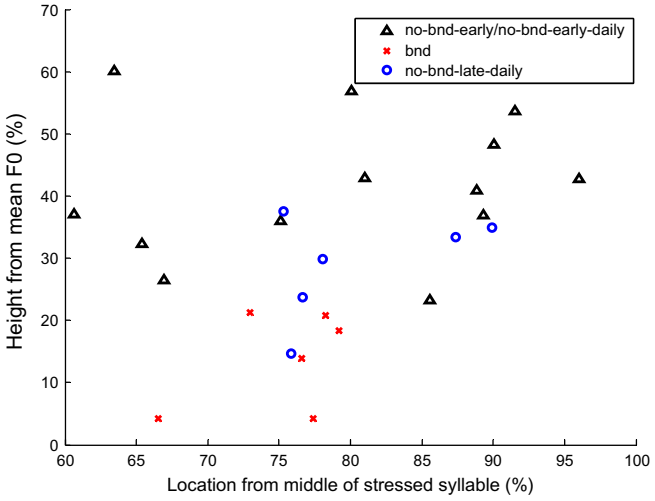

Fig. 12. Scatter plot for the target word *dagada* showing relative F_0 peaks for H* and their relative positions in the accented target vowel for the male speaker M1 in three different contexts: 1. *no-bnd-early*/*no-bnd-early-daily* (triangles); 2. bnd (crosses); 3. no-bnd-late-daily (circles).

observed for 8/10 males and 9/10 females. In this section, the results for the 2 male speakers, denoted by M1 and M2, and the 1 female speaker, denoted by F9, who did not conform to the general trends are discussed.

Fig. 12 shows the scatter plot of the relative F_0 peak heights and their relative positions for the H* accented target word *dagada* for speaker M1. It can be seen that the boundary cases have minimal effect on the peak positions, which appear to be clustered around 75% of the normalized vowel duration. A similar plot was observed for the target word *dada*, with the peak positions appearing around 70%. While speaker M1 did not display any reliable shifts in peak position for any of the three cases, the height of the peak was consistent with general trends; that is, the *bnd* case had a smaller change in height than the *no-bnd-early*(*-daily*) and *no-bnd-late-daily* cases. Thus it appears that this speaker shares some of the more general response to tonal crowding, i.e. a smaller F_0 excursion for the H*, but not others, i.e. an earlier F_0 peak.

Speaker F9 had F_0 peak shifts aligned with the general trend for the target word *dagada*, but no peak movement could be seen for target word *dada*. The opposite was found for the heights of the F_0 peaks, with the target word *dagada* not conforming to the expected trend of having a lower relative height for the *bnd* condition; for this speaker, the F_0 peak height was on average about 30% higher for the *bnd* case than for the other cases (rather than lower, as was the general trend). Again, this speaker showed some of the general responses to tonal crowding, but not as consistently as other speakers, and in at least one measure showed an idiosyncratic response.

Speaker M2 was unlike the majority of the speakers who had clearly detectable F_0 peaks for the H* accented target words, in that he had some F_0 contours whose shape was more similar to a down-stepped (!H*) accent. This occurred in 5/20 and 13/20 utterances for the target words *dagada* and *dada* respectively. Where there was a detectable F_0 peak, this speaker showed no shift of the peak to an earlier location under conditions of crowding for the *dagada* targets, and the relative height of the peak for the *bnd* case was higher than for the other cases.

Interestingly, informal listening showed that very little audible difference could be perceived between these three speakers (M1, M2 and F9) and the others. Table 10 summarizes the inconsistencies of these speakers when compared with the general trends for the *bnd* case, which were: (1) less F_0 peak shift, and (2) smaller F_0 peak change.

On average, the energy measure showed that for cases where the target word was located at the boundary (*bnd*), the normalized energy was lowest, followed by the *no-bnd-late-daily* cases, with the *no-bnd-early(-daily)* cases having the highest energy. This trend was generally adhered to by all of the speakers in the corpus, although for some speakers there was considerable overlap between the *bnd* and the *no-bnd-late-daily* cases; for example, speakers M3, M6 and M9 shown in Fig. 5 for the target word *dagada*. Interestingly, these speakers also displayed the same behavior for the target word *dada*, suggesting perhaps that the energy measure is relatively consistent across the target words for each speaker.

Duration measure patterns were also fairly consistent across the target words for individual speakers. It was hypothesized that the phrase-final syllable would have extra lengthening if it was preceded by a pitch accent, regardless of the type of pitch accent. This hypothesis was shown to be true for all of the speakers if the H* and L* accents were considered together. However, when the pitch accent types were considered separately, as shown in Fig. 11 for female speakers for the target word *dada*, some speakers (F5 and F8) had a shorter duration for the H* preceding case than the no-accent case. These two speakers also showed the same characteristics for the target word *dagada*. Other differences for other speakers were also found to be consistent across target words, suggesting that they are not random variation but controlled choices for parameter values.

While our generalized results provide a very compact way to describe the effects of tonal crowding, the individual differences described in this section show that acoustic correlates of tonal targets can vary substantially between speakers. It would be of interest to explore the possibility that these variations reflect different decisions about which cues to produce, vs. differences in degree of control.

4.3. Theories of tonal crowding

While the effects of tonal crowding have been described in a number of contexts (Arvaniti et al., 2006; Arvaniti and Garding, 2007; Grabe et al., 2000; Ode, 2005), the principles that govern these effects have not been fully and systematically explored. Moreover, tonal crowding effects can be seen as part of the more general question of how F_0 and other acoustic exponents of tonal targets are real-

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

Table 10
Comparison of the speakers M1, M2 and F9's F_0 peak position and relative height consistencies with the general trends for the *bnd* case; a 'Yes'/'No' denotes agreement/disagreement while 'N/A' means no enough data was available.

Speaker	dagada		dada	
	Less F_0 peak shift for *bnd* case	Smaller F_0 peak change for *bnd* case	Less F_0 peak shift for *bnd* case	Smaller F_0 peak change for *bnd* case
M1	No	Yes	No	Yes
M2	No	No	N/A	N/A
F9	Yes	No	No	Yes

ized with respect to the segmental content of an utterance. A number of issues are raised by earlier work in this area, including the following:

(1) *What are the options for a speaker when confronted by the need to realize several tonal targets in quick succession*? Several different mechanisms have been proposed, including truncation (i.e. a smaller F_0 movement, Grabe et al. (2000)), compression (i.e. a faster F_0 movement, Fougeron and Jun (1998)), and deletion (i.e. elimination of one of the tonal targets, Levi (submitted for publication) for Turkish; Fougeron and Jun for French). Another possible response is to move one or both of the two target realizations within their syllables so they occur further apart; articulating an F_0 movement earlier in its syllable would be one response of this type. Yet another possibility is to lengthen the segmental material, particularly the syllabic nucleus, i.e. slowing the speaking rate (at least temporarily) to make time for the realization of complex tone sequences. It appears that most of the speakers in this experiment, when confronted by tonal crowding from a pitch accent and boundary-related tones on the final two syllables of an utterance, chose a combination of a less-extreme F_0 movement, a longer duration of the syllable and earlier realization of the F_0 peak. The choices that speakers make among these possible alignment adjustment mechanisms, and the details of how those choices are realized, are in need of further investigation. For example, when two targets move apart in response to crowding, does just one of them move to an earlier location, or do both tonal targets move away from each other?

(2) *What is the definition of crowding*? That is, how close to each other do two tonal targets have to be, in order to influence each other's realization, and how is this distance measured? Is it in terms of time increments, i.e. milliseconds? In terms of the number of voiced phonological segments that are available to carry the F_0? In terms of constituent structure (e.g. syllables)? Or in some combination of these scales? It has been suggested that, in order to eliminate the effect of one tonal target on another, it may be necessary to have two unstressed syllables between the two targeted syllables; others have suggested that the preferred target-syllable relationship is one target per syllable (Arvaniti et al., 2006; Levi, submitted for publication). This decision is important because its result will inform the estimation of the preferred, non-crowded realization of a tonal target sequence, as well as the generation of appropriate algorithms for natural-sounding synthesis.

(3) *Do different tonal target types respond differently to crowding*? That is, do all types of tonal targets follow the same principles of interaction? For example, do two adjacent pitch accents interact in the same way as a pitch accent followed by a boundary tone? Do bi-tonal targets behave differently from single tonal targets. High targets differently from Lows? Our results suggest that L^* targets behave differently in response to tonal crowding than H^* targets, and Silverman and Pierrehumbert (1990) results suggest that pre-nuclear and nuclear accents respond in largely similar ways. But the full scope of interactions among various tonal target types has not been investigated.

(4) *Do different languages and dialects exhibit different principles of interaction*? For example, Grabe et al. (2000) reported substantial differences in tonal target interactions in various dialects of British English. Similarly, Mücke and her colleagues report differences between Viennese and Dusseldorf German for resolution of such alignment issues (Mücke et al., 2009), and Levi (submitted for publication) reports target deletion by Turkish speakers to prevent tonal crowding.

While we do not yet have a clear picture of how these issues are resolved in American English speech, we can test the hypothesis that F_0 peaks seem to occur earlier in syllables that are lengthened phrase-finally, i.e. if the peak occurs at a fixed number of milliseconds from the V onset, it will seem to occur earlier in longer syllables. If final lengthening were responsible for the early H^* peaks, then in our *bnd* condition the H^* peak locations in the raw vowel durations should show one cluster of points for all three cases (early, late, boundary). Results are shown in Fig. 13 which plots the time of the peak in milliseconds from the V onset against the height of the peak above the mean F_0, for two typical speakers, one male for target word *dagada* and one female for target word *dada*. For most speakers the peak position was significantly earlier for the *bnd* case, as per our results for normalized vowel durations, showing that the peak occurs earlier in absolute as well as relative terms under tonal crowding.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005

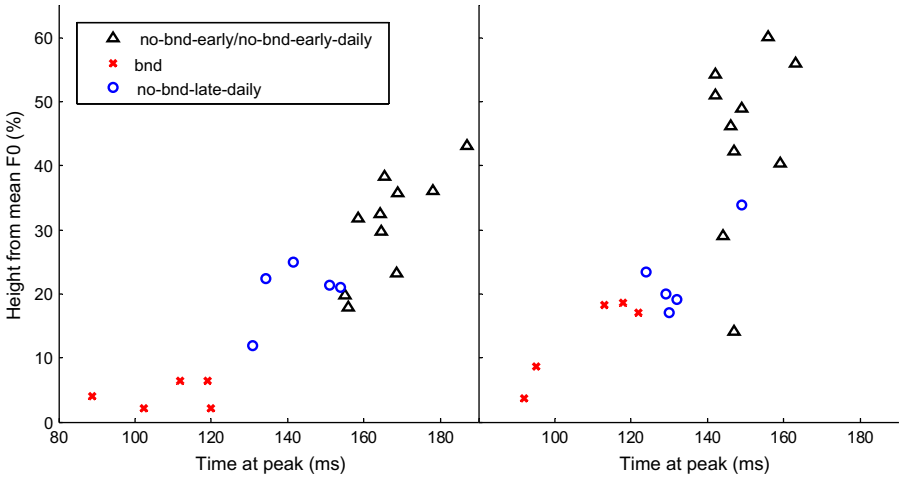

Fig. 13. Scatter plot showing the times for the raw peak position in milliseconds from the V onset and the normalized peak heights for the three cases; *no-bnd-early*(*-daily*) (triangles), *bnd* (crosses) and *no-bnd-late-daily* (circles). The left panel shows the times and heights for a typical male speaker for the target word *dagada* and the right panel shows the results for a typical female speaker for the target word *dada*.

As Silverman and Pierrehumbert (1990) point out, the question of how tonal crowding affects the phonetic realization of pitch accents and other tonal targets is part of the larger general issue of tonal alignment. With this in mind, they evaluated several alternative mechanisms for differences in F_0 peak alignment across stimulus types, including *invariant duration* of the F_0 rise; *gestural overlap* and truncation; *tonal repulsion* with earlier gesture beginning; *phonological mediation* by the addition of extra beats to the metrical grid, lengthening the syllable with the result that the F_0 peak occurs earlier; and *sonority profile*.

In summary, although a clear picture of the factors that govern tonal alignment for different tonal target types in various contexts and in different languages is still emerging, studies such as this one have begun to reveal some aspects of these patterns for individual languages. It is clear that considerable further research is needed to clarify the factors that govern the alignment of F_0 contours with the words and syllables of a spoken utterance, and how the alignment changes under conditions of tonal crowding.

5. Conclusion

This study compares the acoustic characteristics of two types of pitch accents in American English (H^* and L^*) in three types of locations within the phrase (in early and late non-phrase-final words and phrase-final words), in two types of target words (two- and three-syllable words) with penultimate lexical stress. Results for F_0 show that for most speakers and for both target words, the F_0 peak for a nuclear H^* accent occurs earlier in conditions of tonal crowding due to phrase-final boundary tones in the same word, and is realized with a lower F_0. In contrast, F_0 troughs for L^* accents did not show the same effects of tonal crowding, suggesting that H^* and L^* accents may not be realized according to the same principles. Comparison across 20 individual speakers (10 male and 10 female) revealed that the general findings are robust, but that three speakers did not comport with all aspects of the general findings, raising the possibility that some speakers may employ idiosyncratic cue patterns. Analysis also showed that energy levels decrease across the utterance, and that a phrase-final syllable is longer if the immediately-preceding syllable in the final word is accented than if it is not. Taken together, these results highlight the importance of taking context into account for prosodic analysis.

Acknowledgements

We thank Dr. Patricia Keating for her helpful suggestions and advice during the preparation of this study. We also thank the speakers who participated in this experiment. This work was supported in part by the NSF.

References

Arvaniti, A., Garding, G., 2007. Dialectal variation in the rising accents of American English. In: Cole, J., Hualde, J.H. (Eds.), Papers in Laboratory Phonology 9, pp. 547–576.

Arvaniti, A., Ladd, D.R., Mennen, I., 2006. Phonetic effects of focus and "tonal crowding" in intonation: evidence from Greek polar questions. Speech Commun. 48, 667–696.

Beckman, M., Edwards, J., 1994. Articulatory evidence for differentiation stress categories. Lab. Phon. III., 7–33.

Beckman, M., Pierrehumbert, J.B., 1986. Intonational structure in Japanese and English. Phonol. Yearbook 3, 255–309.

Chafe, W.L., 1993. Prosodic and functional units of language. In: Edwards, J.A., Lampert, M.D. (Eds.), Talking Data: Transcription and Coding in Discourse Research. Lawrence Erlbaum, Hillsdale, NJ, pp. 3–43.

Dilley, L., Shattuck-Hufnagel, S., Ostendorf, M., 1996. Glottalization of vowel-initial syllables as a function of prosodic structure. J. Phonetics 24, 423–444.

Fougeron, C., Jun, S.-A., 1998. Rate effects on French intonation: prosodic organization and phonetic realization. J. Phonetics 26, 45–69.

Grabe, E., Post, B., Nolan, F., Farrar, K., 2000. Pitch accent realization in four varieties of British English. J. Phonetics 28, 161–185.

Hirschberg, J., Pierrehumbert, J., 1986. The intonational structuring of discourse. In: Proc. 24th Annual Meeting on Association for Computational Linguistics, pp. 136–144.

Iseli, M., Shue, Y.-L., Alwan, A., 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. J. Acoust. Soc. Am. 121 (4), 2283–2295.

Jilka, M., Möbius, B., 2007. The influence of vowel quality features on peak alignment. In: Proc. Interspeech, Antwerp, Belgium, pp. 2621–2624.

Kawahara, H., de Cheveigné, A., Patterson, R.D., 1998. An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMP in the STRAIGHT-suite. In: Proc. ICSLP, Sydney, Australia.

Klatt, D.H., 1976a. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. J. Acoust. Soc. Am. 59, 1208–1221.

Klatt, D.H., 1976b. Structure of a phonological rule component for a synthesis-by-rule program. IEEE Trans. Acoust. Speech Signal Process. 35 (4), 445–472.

Kochanski, G., Grabe, E., Coleman, J., Rosner, B., 2005. Loudness predicts prominence: fundamental frequency lends little. J. Acoust. Soc. Am. 118 (2), 1038–1054.

Ladd, D. Robert, 1996/2008. Intonational Phonology. Cambridge University Press, Cambridge.

Levi, S.V., submitted for publication. Intonation in Turkish: the realization of noun compounds and genitive possessive NPs.

Mücke, D., Grice, M., Becker, J., Hermes, A., 2009. Sources of variation in tonal alignment: evidence from acoustic and kinematic data. J. Phonetics. doi:10.1016/j.wocn.2009.03.005.

Ode, C., 2005. Neutralization or truncation? The perception of two Russian pitch accents on utterance-final syllables. Speech Commun. 47, 71–79.

Ohala, J.J., Ewan, W.G., 1973. Speed of pitch change (A). J. Acoust. Soc. Am. 53 (1), 345.

Pierrehumbert, J.B., 1980. The phonology and phonetics of English intonation. Dissertation, MIT.

Pierrehumbert, J.B., Talkin, D., 1991. Lenition of /h/ and glottal stop. In: Papers in Laboratory Phonology II. Cambridge University Press, Cambridge, UK, pp. 90–117.

Rosenberg, A., Hirschberg, J., 2006. On the correlation of energy and pitch accent in read English speech. In: Proc. Interspeech, Pittsburgh, pp. 301–304.

Shattuck-Hufnagel, S., Turk, A., 1996. A prosody tutorial for investigators of auditory sentence processing. J. Psycholinguist. Res. 25 (2), 193–247.

Shue, Y.-L., Iseli, M., Veilleux, N., Alwan, A., 2007. Pitch accent versus lexical stress: quantifying acoustic measures related to the voice source. In: Proc Interspeech, Antwerp, Belgium, pp. 2625–2628.

Shue, Y.-L., Shattuck-Hufnagel, S., Iseli, M., Jun, S., Veilleux, N., Alwan, A., 2008. Effects of intonational phrase boundaries on pitch-accented syllables in American English. In: Proc. Interspeech, Brisbane, Australia, pp. 873–876.

Silverman, K.E.A., Pierrehumbert, J.B., 1990. The timing of prenuclear high accents in English. In: Kingston, J., Beckman, M.E. (Eds.), Papers in Laboratory Phonology 1: Between the Grammar and the Physics of Speech. Cambridge University Press, Cambridge UK, pp. 72–106.

Slifka, J., 2007. Some physiological correlates to regular and irregular phonation at the end of an utterance. J. Voice 20, 171–186.

Sluijter, A.M.C., van Heuvan, V.J., 1996a. Spectral balance as an acoustic correlate of linguistic stress. J. Acoust. Soc. Am. 100 (4), 2471–2485.

Sluijter, A.M.C., van Heuvan, V.J., 1996b. Acoustic correlates of linguistic stress and accent in Dutch and American English. In: Proc. ICSLP, Philadelphia, PA, pp. 630–633.

Sundberg, J., 1979. Maximum speed of pitch changes in singers and untrained subjects. J. Phonetics 7, 71–79.

Titze, I.R., 1989. Physiologic and acoustic differences between male and female voices. J. Acoust. Soc. Am. 85 (4), 1699–1707.

Turk, A.E., Shattuck-Hufnagel, S., 2007. Phrase-final lengthening in American English. J. Phonetics 35 (4), 445–472.

Turk, A.E., White, L., 2007. Structural effects on pitch accentual lengthening in English. J. Phonetics 27, 171–206.

Xu, Y., Sun, X., 2002. Maximum speed of pitch change and how it may relate to speech. J. Acoust. Soc. Am. 111 (3), 1399–1413.

Please cite this article in press as: Shue, Y.-L. et al., On the acoustic correlates of high and low nuclear pitch accents in American English, Speech Comm. (2009), doi:10.1016/j.specom.2009.08.005