Pitch pattern variations in three regional varieties of American English

Robert Allen Fox, Ewa Jacewicz, Jessica Hart

Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, USA.

`fox.2@osu.edu, jacewicz.1@osu.edu, hart.439@buckeyemail.osu.edu`

Abstract

This acoustic study explored dialect effects on realization of nuclear pitch accents in three regional varieties of American English spoken in central Ohio, southeastern Wisconsin and western North Carolina. Fundamental frequency (f0) change from vowel onset to offset in the most prominent syllable in a sentence was examined along four parameters: maximum f0 change, relative location of f0 maximum, f0 offset and f0 fall from maximum to offset. A robust finding was that the f0 contours in the Southern (North Carolina) variants were significantly distinct from the two Midwestern varieties whose contours did not differ significantly from one another. The Southern vowels had an earlier f0 rise, a greater f0 fall and a lower f0 offset than either Ohio or Wisconsin vowels. There was a sharper f0 drop preceding a voiceless than a voiced syllable coda. No significant dialect-related differences were found for flat f0 contours in unstressed vowels, which were also examined in the study. This study contributes the finding that dynamic variations in pitch are greater for vowels which also exhibit a greater amount of spectral dynamics. The interaction of these two sets of cues contributes to the melodic component associated with a specific regional accent.

Index Terms: pitch accent, f0 variation, regional accent, American English.

1. Introduction

The traditional view of the English vowel system is that monophthongs are static sounds and diphthongs are dynamic in nature. The existence of transitional spectral changes in diphthongs is widely recognized as these changes can cue phonemic contrasts (e.g., *high–how*). It is less well recognized that spectral changes also occur in nominal monophthongs although their function is clearly subphonemic. Vowel inherent spectral change (VISC)–the change in spectral properties over the time course of a vowel–was found in American, Canadian, and Australian English monophthongal vowels more than two decades ago [1, 2, 3]. However, with the advancement of speech technology, this vowel feature is currently receiving a wider attention in theoretical and applied research [4]. For example, is has been shown that dynamic spectral structure is necessary for optimal vowel recognition [5] and that non-native speakers must acquire relevant information in VISC in order to achieve native-like levels of performance [6].

Recent investigations of vowel systems in several regional varieties of American English brought to light that the variations in dynamic spectral characteristics are systematic and that these variations differentiate dialects and generations of speakers across regions in the United States [7, 8, 9]. Dialectal variants of the same vowel category were found to differ in the nature and extent of VISC, indicating that this vowel feature is important to speakers of regional varieties because it has been transmitted from one generation to another [10].

In this study, we explore another dynamic property of vowels related to the characteristics of the source (and not the filter). We examine a distinctive use of pitch over the time course of a vowel in three different regional varieties of American English. The possibility that different dialects use distinctive pitch accents in vowels was explored in several dialects of English spoken in the British Isles as a part of a larger investigation of dialect variation in intonation patterns [11]. Although cross-dialectal differences were found, there was also a considerable overlap among the dialects in the types of the nuclear accents produced. Nevertheless, differences in pitch accent realization in British English do exist and have been manifested along several variables [12]. However, the dialect-related prosodic differences have been demonstrated primarily within larger phrasal domains, including the alignment of tones with syllables and phrase boundaries [13, 14]. Recently, significant effects of regional dialect were also found in American English, including the distributions of pauses, pitch accents, and phrasal–boundary tone combinations [15].

Our interests in this paper are in the dialect-specific use of fundamental frequency (f0) contours in American English to convey nuclear accents. We hypothesize that there is a meaningful interaction between melodic contour in a vowel related to pitch movement and VISC. The distinctive pitch patterns (pitch glides) interact with distinctive spectral changes (VISC) and this coordinated use of dynamic source and spectral characteristics creates vowel-specific "signature tune" associated with a regional accent. While modeling of this complex interaction awaits future investigation, we explore here the dynamics of a pitch accent (i.e., f0 change from vowel onset to offset) in the most prominent syllable in the phrase along four parameters: maximum of f0 change, relative location of f0 maximum, f0 offset and f0 fall from maximum to offset. In addition to nuclear accent, we also examine whether dialect-related differences in pitch contour exist in the absence of prosodic prominence, i.e., in unstressed syllables.

2. Methods

The f0 patterns were examined in three regional dialects of American English. We selected two Midwestern varieties: A Midland variety spoken in central Ohio (OH) and a variety spoken in southeastern Wisconsin (WI), in the dialect region classified as Inland North [16]. As a third variety, we selected a Southern dialect spoken in the Appalachian region in western North Carolina (NC), classified as Inland South. There is much anecdotal evidence that Southern English, in general, and Appalachian English, in particular, is "more expressive" than are the other two regional varieties [17]. One source of this expressiveness is related to distinct patterns of VISC in this dialect [8]. We examine here whether notable differences also exist in f0 contour shape which may contribute to the impression of greater melodic variations in Appalachian English.

2.1. Speakers

24 women aged 50-64 years old, 8 for each dialect variety, produced the speech samples. Defined geographically, these speakers created highly homogeneous samples of regional speech because they were born, raised and spent most of their lives in one of the three narrow areas: Columbus and suburbs (Ohio), Madison and suburbs (Wisconsin) and the Jackson county area (North Carolina). None of the speakers reported any speech disorders.

2.2. Speech materials and procedure

Five vowels differing in the amount of VISC (or a degree of diphthongization) were selected: /ɪ, ɛ, e, æ, aɪ/. Each vowel was contained in a monosyllabic word of the structure /b_dz/ and /b_ts/ (before voiced and voiceless coda, respectively) and produced in a sentence. The sentences were constructed to elicit: 1) the nuclear accent on the most prominent syllable corresponding to the main sentence stress, and 2) a low prosodic prominence corresponding to unstressed position in a sentence.

Examples (nuclear accent (in bold)):

*Ted thinks the fall **BIDS** are low.*
*Rob said the tall **BEDS** are warm.* (1)
*Mike thinks the small **BADS** are worse.*

Examples (unstressed position (in bold)):

*Ted thinks the fall **bids** are LOW.*
*Rob said the tall **beds** are WARM.* (2)
*Mike thinks the small **bads** are WORSE.*

Recordings were controlled by a custom program in Matlab which displayed a sentence to be read by the speaker on the computer monitor. The sentences were presented in random order. A head-mounted Shure SM10A dynamic microphone was used, positioned about 1.5 in. from the speaker's mouth. The samples were recorded and digitized at a 44.1-kHz sampling rate with 16-bit quantization. The speaker read the sentence placing the main sentence stress on the word in all caps. Only fluent productions (without pauses) were accepted. For that reason, multiple repetitions of each sentence were obtained (as many as needed) to select the three most fluent repetitions for subsequent acoustic analysis. A total of 1408 sentences were analyzed, 60 sentences from each speaker (except for one speaker who produced 30 sentences).

2.3. f0 measurements

Vowel onsets and offsets for all target vowels identified using a waveform editing program (Audobe Audition) and checked (by two different researchers) using a custom Matlab program that displayed the target word, target vowel and marked word and vowel onsets and offsets.

After these landmark locations had been identified, f0 measurements were made using a different group of custom Matlab programs. Overall f0 was computed using autocorrelation analysis over the entire duration of the vowel. Next, f0 autocorrelation measurements were made in a series of 16 ms windows (with 50% overlap) over the course of the vowel. Following these measurements, another program displayed both the overall and individual segment f0 values and allowed hand correction of mistracked f0 values (using TF32 [18]). These hand-corrections were then checked (and modified where deemed necessary) by one of the authors. All measurements were then time-normalized to a 0-100 point scale (based on the time proportions for each separate vowel) with f0 values between actual measurement points based on linear interpolation.

Given differences in basic speaking f0s among speakers (related to a number of physiological features including size of the vocal folds), examination of the prosodic "melody" of the vowel (which may be linked to linguistic properties [19]) on the basis of the original Hz measurements would be hampered by such variation. Therefore, in this study we examine the changes in f0 relative to the onset frequency using the semitone scale (in terms of cents, which is 1/100 of a semitone)—this scale also more appropriately reflect speakers' (and listeners') intuition regarding intonational spans across speakers [20]. The time-normalized f0 change values (at normalized time points from n=0 to 100) were converted to cents using the following formula:

$$f0\_change_n = 1200 * \log_2 (f0_n \ / f0_0) \qquad (3)$$

where $f0_0$ represents the frequency of f0 at vowel onset

Figure 1 shows the schematic of four measurements used in the study: maximum of f0 change (max value of f0 change), relative location of f0 maximum (time when max f0 value occurs), f0 offset (f0 change value at offset) and amount of f0 decrement (f0 change from max to offset).

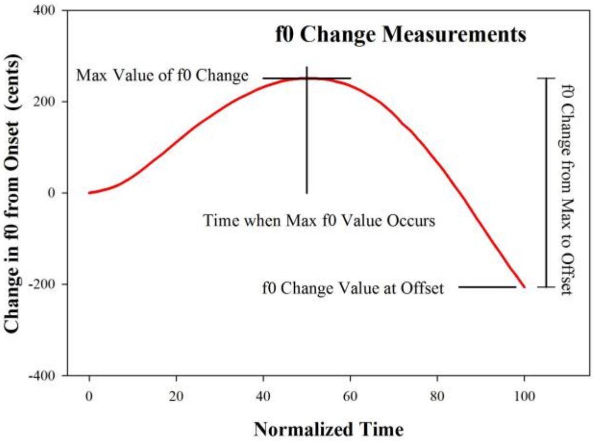

Figure 1: *Schematic of four f0 measurements.*

2.4. Results

Prior to examination of f0 change patterns, we first considered the variation in the speaking f0 (or, in this case, reading f0) across speakers of different dialects. Table 1 summarizes the overall f0 means for vowels before voiceless and voiced coda for stressed and unstressed variants. A repeated-measures ANOVA with the within-subject factors stress level and coda type and dialect as a between-subject factor was used to determine any significant differences. There was a significant effect of stress level ($F(2,21)=135.9$, $p<.001$, $\eta^2=.866$). Stressed vowels (241.9 Hz) had a higher overall f0 than did unstressed vowels (170.3 Hz). However, no other main effects or interactions were significant, indicating that neither speaker dialect nor voicing status of the syllable coda appreciably affected the overall speaking f0.

Table 1. *Mean (s.e.) overall f0 values (in Hz) for Ohio (OH), Wisconsin (WI) and North Carolina (NC) speakers*.

Coda	OH	WI	NC	Total
Voiceless				
Stressed	253.6	235.8	235.3	242.5
	(13.8)	(9.2)	(13.7)	(7.1)
Unstressed	187.4	159.2	164.2	170.3
	(10.3)	(5.8)	(7.1)	(5.1)
Voiced				
Stressed	252.9	235.8	235.1	241.2
	(15.1)	(9.6)	(13.7)	(7.4)
Unstressed	184.9	161.1	164.8	170.3
	(10.6)	(6.0)	(6.1)	(4.9)
Totals				
Stressed	253.2	237.3	235.2	241.9
	(9.9)	(6.5)	(9.4)	(5.1)
Unstressed	186.2	160.1	164.5	170.3
	(7.2)	(4.0)	(4.5)	(3.5)

Next, the changes in f0 contours collapsed across vowels were examined and assessed by separate repeated-measures ANOVAs for each stress level and each coda type along the four parameters: f0 max, relative location of f0 max, f0 offset, and f0 change (from max to offset).

2.4.1. *Stressed vowels before voiceless coda*

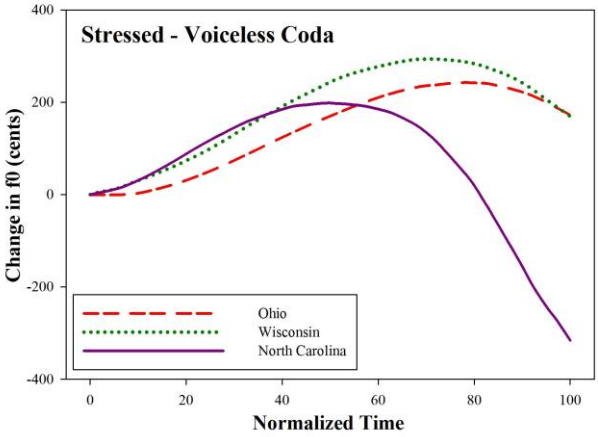

Figure 2: *Mean f0 contour for stressed vowels in /b_ts/.*

Figure 2 shows the mean smoothed spline f0 change contours collapsed across the vowels produced by the three groups of speakers in stressed positions before a voiceless coda. The analyses indicated that f0 max value did not differ significantly as a function of dialect. However, significant dialect differences were found for the three remaining variables. For the relative location of f0 max ($F(2,21)=8.83$, $p=.002$, $\eta^2=.457$), NC f0 max occurred earlier in time than WI or OH, which did not differ significantly one from another (Scheffe at .05 level). For the f0 offset ($F\ (2,21)=6.27$, $p=.001$, $\eta^2=.374$), NC value was significantly lower than both OH and WI, which also did not differ one from another. For f0 change from max to offset ($F(2,21)=16.9$, $p<.001$, $\eta^2=.617$), NC had a significantly greater drop than either OH or WI, which again were not different. Clearly, NC vowels differed from the two dialects in that they had an earlier rise (f0 max occurred earlier in time), a greater fall and a lower offset whereas both the OH and WI variants had a late rise and a very small fall.

2.4.2. *Stressed vowels before voiced coda*

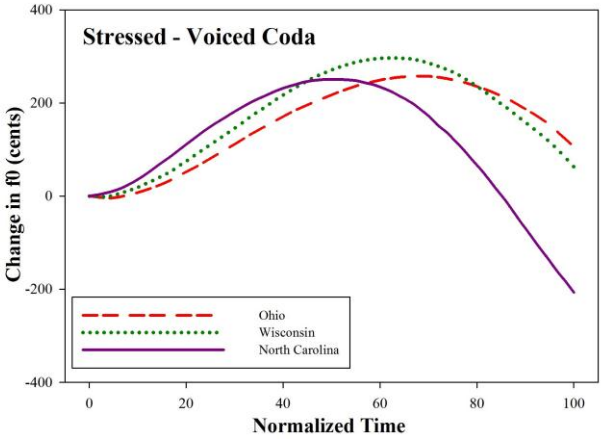

Figure 3: *Mean f0 contour for stressed vowels in /b_dz/.*

Figure 3 shows the corresponding f0 contours for the vowels in stressed positions before a voiced coda. Although the difference between the NC contour and the contours of either OH or WI was comparatively less dramatic, statistical analyses yielded identical results. There were no significant dialect-related differences in f0 max value for vowels before voiced coda. NC f0 max again occurred significantly earlier in time than either OH or WI ($F(2,21)=7.46$, $p=.004$, $\eta^2=.415$), NC f0 offset was significantly lower than either OH and WI ($F\ (2,21)=5.22$, $p=.014$, $\eta^2=.332$), and the f0 drop was larger for NC than for either OH or WI ($F(2,21)=6.97$, $p=.005$, $\eta^2=.399$). We note, however, that the effect size (η^2) for the f0 drop was smaller for the vowels before a voiced coda compared to the vowels before a voiceless coda. This was primarily due to a comparatively higher NC f0 offset and lower OH and WI offset values, which reduced the larger dialect differences found before the voiceless coda.

2.4.3. *Unstressed vowels*

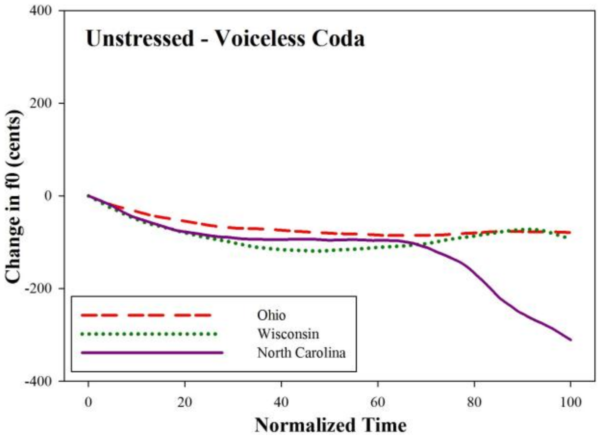

Figure 4: *Mean f0 contour for unstressed vowels in /b_ts/.*

Figures 4 and 5 show the f0 contours for the vowels in unstressed positions. As can be seen, the contours are flat and there are no dialect-related differences except for the relatively small f0 drop for NC vowels before the voiceless coda. However, as a whole, the results for the unstressed vowels indicate that significant dialectal differences in pitch movement may arise when the vowels convey nuclear accents but not when the associated pitch movement is absent.

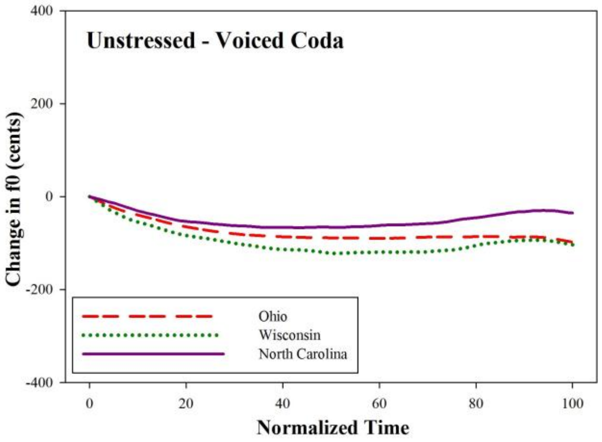

Figure 5: *Mean f0 contour for unstressed vowels in /b_dz/.*

3. Discussion

This study explored dialect effects on realization of nuclear pitch accents in three regional varieties of American English. We selected two relatively close Midwestern varieties in terms of their shared vowel features and temporal characteristics (OH and WI) and a distant Southern Appalachian variant (NC), whose vowels exhibit greater spectral changes and longer durations. We also examined whether there were any dialectal differences in the production of unstressed vowels in these dialects. Due to the preliminary nature of this study, the f0 variation in vowels was assessed in more global terms, by collapsing measurements for five distinct vowel categories. Variation for individual vowels was not explored at present.

A robust finding was that the f0 contours in the Southern NC variants were significantly distinct from the two Midwestern varieties whose contours did not differ significantly from one another. NC vowels had an earlier f0 rise (in normalized time), a greater f0 fall and a lower f0 offset than either OH or WI vowels. There was a sharper f0 drop – and thus the contour was more "exaggerated" – in NC vowels preceding a voiceless coda than a voiced coda, which can be attributable to a more rapid change in sound energy as a function of vowel shortening (changes in duration of these vowels as a function of stress were explored in [8]).

Notably, there were no dialectal differences for the measure f0 max, indicating that the current speakers produced f0 peaks in a similar manner. The dialect-related differences only became apparent when looking at the other three variables, including time when the peak occurred and the nature of the f0 fall.

The present results for unstressed vowels revealed no dialectal differences for flat f0 contours. Thus, any dialect effects on the shape of f0 contour are manifested when this contour reflects pitch accent. Dialect-related variation pertains primarily to the realization of pitch accents and no systematic variation in f0 occurs when syllables are unstressed.

This study was undertaken as a first step toward testing the hypothesis that there is a meaningful interaction of dynamic cues in vowels related to both spectral changes and variation in pitch. We hypothesize that a combination of these two sets of cues contributes to the melodic aspects of regional accents which act as dialect-specific "signature tunes." The current results for the Southern NC vowels are encouraging, indicating that the greater f0 contour shape interacts with a greater dynamic formant movement. While a more nuanced understanding of this interaction and its effects on the listener await future exploration in structured perception experiments, the contribution of this study is the finding that the Southern pitch contours are more "melodic" in nature relative to the pitch patterns in Midwestern vowels. It is possible, however, that similar interaction of dynamic cues comes into play for selected individual Midwestern vowels, which is not reflected in the overall mean values examined here. This possibility also awaits future exploration.

4. Conclusions

This study provided acoustic evidence that regional dialects of American English may differentially use variation in f0 contour shape to convey nuclear accent in vowels. Overall, the dynamic changes in f0 contours of Southern American vowels were significantly greater than the corresponding pitch patterns in Midwestern vowels. These results supplement earlier findings with respect to the variation in spectral dynamics of Southern versus Midwestern vowels. In particular, there seems to be a correspondence between the greater amount of spectral changes (VISC) and exaggerated pitch contours and the interaction of these acoustic cues may produce a distinct dialect-specific type of melodic variation associated with Appalachian English accent. Conversely, the less dramatic pitch changes correspond to the smaller spectral changes in Midwestern vowels, and this association may denote dialect characteristics typical of the Midwest. This will be explored in greater detail in future work.

5. Acknowledgements

Parts of this work were supported by NIDCD/NIH research grant R01DC006871.

6. References

[1] Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K., "Effects of consonantal environment on vowel formant patterns", Journal of the Acoustical Society of America, 97: 3099-3111, 1995.

[2] Andruski, J. E., and Nearey, T. M., "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables", Journal of the Acoustical Society of America, 91: 390-410, 1992.

[3] Watson, C. I., and Harrington, J., "Acoustic evidence for dynamic formant trajectories in Australian English vowels", Journal of the Acoustical Society of America, 106: 458-468, 1999.

[4] Morrison, G. S., and Assmann, P. F. [Eds], "Vowel inherent spectral change", Springer, 2013.

[5] Hillenbrand, J. M., "Static and dynamic approaches to vowel perception", in G. S. Morrison and P. F. Assmann [Eds], Vowel inherent spectral change, 9-30, Springer, 2013.

[6] Rogers, C. L., Glasbrenner, M. G., DeMasi, T. M., and Bianchi, M., “Vowel inherent spectral change and the second-language learner”, in G. S. Morrison and P. F. Assmann [Eds], Vowel inherent spectral change, 231-259, Springer, 2013.

[7] Jacewicz, E., Fox, R. A., and Salmons, J., “Cross-generational vowel change in American English”, Language Variation and Change, 23:45-86, 2011.

[8] Fox, R., A., and Jacewicz, E., “Cross-dialectal variation in formant dynamics of American English vowels”, Journal of the Acoustical Society of America, 126: 2603-2618, 2009.

[9] Jacewicz, E., Fox, R. A., and Salmons, J., “Vowel change across three age groups of speakers in three regional varieties of American English”, Journal of Phonetics, 39: 683-693, 2011.

[10] Jacewicz, E., Fox, R. A., and Salmons, J., “Regional dialect variation in the vowel systems of typically developing children”, Journal of Speech, Language, and Hearing Research, 54: 448-470, 2011.

[11] Grabe, E., “Intonational variation in urban dialects of English spoken in the British Isles”, in P. Gilles, and J. Peters [Eds], Regional variation in intonation, 9-31, Linguistische Arbeiten, Tuebingen, Niemeyer, 2004.

[12] Grabe, E., Post, B., Nolan, F., and Farrar, K., “Pitch accent realization in four varieties of British English”, Journal of Phonetics, 28: 161-185, 2000.

[13] Grabe, E., and Post, B., “Intonational variation in the British Isles”, in B. Bel, and I. Marlin [Eds], Proceedings of Speech Prosody 2002, 343-346, Aix-en-Provence, France, Laboratoire Parole et Langue, 2002.

[14] Arvaniti, A, and Garding, G., “Dialectal variation in the rising accents of American English”, in J. Cole, and J. Hualde [Eds], Laboratory Phonology, 9: 547-576, Mouton de Gruyter, 2007.

[15] Clopper, C. G., and Smiljanic, R., “Effects of gender and regional dialect on prosodic patterns in American English”, Journal of Phonetics, 39: 237-245, 2011.

[16] Labov, W., Ash, S., and Boberg, C., “Atlas of North American English: Phonetics, phonology and sound change”, Mouton de Gruyter, 2006.

[17] Wolfram, W., and Christian, D., “Appalachian speech”, Arlington: Center for Applied Linguistics, 1976.

[18] Milenkovic, P., “TF32 software program”, University of Wisconsin, Madison, 2003.

[19] Ladd, D. R., “Intonational phonology”, 2nd ed., Cambridge University Press, 2008

[20] Nolan, F., “Intonational equivalence: an experimental evaluation of pitch scales”, Proceedings of the 15th International Congress of Phonetic Science, Barcelona, Spain, 774-777, 2003.