

# Cross-dialectal variation in formant dynamics of American English vowels

Robert Allen Fox<sup>a)</sup> and Ewa Jacewicz

*Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210-1002*

(Received 31 March 2009; revised 30 July 2009; accepted 30 July 2009)

This study aims to characterize the nature of the dynamic spectral change in vowels in three distinct regional varieties of American English spoken in the Western North Carolina, in Central Ohio, and in Southern Wisconsin. The vowels /ɪ, ε, e, æ, aɪ/ were produced by 48 women for a total of 1920 utterances and were contained in words of the structure /bVts/ and /bVdz/ in sentences which elicited nonemphatic and emphatic vowels. Measurements made at the vowel target (i.e., the central 60% of the vowel) produced a set of acoustic parameters which included position and movement in the F1 by F2 space, vowel duration, amount of spectral change [measured as vector length (VL) and trajectory length (TL)], and spectral rate of change. Results revealed expected variation in formant dynamics as a function of phonetic factors (vowel emphasis and consonantal context). However, for each vowel and for each measure employed, dialect was a strong source of variation in vowel-inherent spectral change. In general, the dialect-specific nature and amount of spectral change can be characterized quite effectively by position and movement in the F1 by F2 space, vowel duration, TL (but not VL which underestimates formant movement), and spectral rate of change.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3212921]

PACS number(s): 43.70.Fq, 43.72.Ar [DAB]

Pages: 2603–2618

## I. INTRODUCTION

This study is an acoustic investigation into time-varying spectral features of the vowel target and the use of these features by different regional variants (dialects) of American English. In the long tradition of research on vowel acoustics, the vowel target has been regarded as the central section of the vowel which is relatively unaffected by surrounding consonants (e.g., [Lehiste and Peterson, 1961](#); [Lindblom, 1963](#)). As such, formant frequencies measured at the vowel target are often considered to characterize most appropriately a particular vowel quality with the general assumption that this target implies some degree of invariance or vowel's "steady state." As the research progressed, however, it has been recognized that even the most monophthongal vowel is never truly "static" and usually exhibits a certain amount of vowel-inherent spectral change (e.g., [Harrington and Cassidy, 1994](#); [Nearey and Assmann, 1986](#)). This spectral change is not contextually determined but is a systematic property of the vowel itself.

Naturally, the presence of dynamic spectral change in a vowel gave rise to investigation in the perceptual domain, inviting the question of how important are the dynamic cues to vowel identification. Certainly a substantial body of research has focused on determining which aspects of the acoustic signal contribute most to vowel identification: a relatively steady target or rapid consonantal transitions. For example, work by Strange and Jenkins (e.g., [Jenkins \*et al.\*, 1983, 1994](#); [Strange, 1987, 1989](#); [Strange \*et al.\*, 1976, 1983](#)) underscored the role of consonantal transitions proposing that a vowel can be reliably identified even if its "center" has

been removed experimentally from the signal. On the other hand, studies by Nearey and colleagues (e.g., [Hillenbrand and Gayvert, 1993](#); [Hillenbrand and Nearey, 1999](#); [Nearey and Assmann, 1986](#); [Kewley-Port and Neel, 2006](#)) demonstrated that listeners identified vowels with greater accuracy when the vowel-specific pattern of spectral change was preserved at the vowel's center. The results from these two lines of research suggest that neither the vowel target nor consonantal transitions alone are fully sufficient for vowel identification.

The complexity of acoustic variation in formant dynamics comes from several sources, including the vowel-specific nature of trajectory change in the formant space, consonantal context effects, emphatic stress, or broadly defined prosodic effects. The difficulty in characterizing the dynamic spectral changes throughout the course of the vowel lies in the fact that these changes occur in time and are subject to temporal variation in speech. That is, the amount of vowel-inherent spectral change may vary with vowel duration, which is also affected systematically by consonantal and prosodic contexts as well as by variation in speech tempo. For example, consonantal context effects may persist throughout the vowel (including its target) if the vowel is sufficiently short. This leads to yet another complication, however, in that vowel target may not be easily defined as it is the entire formant trajectory that undergoes changes over time.

The presence of time-varying spectral features such as the amount of spectral change and spectral rate of change (roc) implies that the dynamic information includes cues from both spectral and temporal domains. As it has been shown and is argued below, both sets of cues are important to a better understanding of vowel dynamics and their use in speech communication. For example, according to the contextual model of articulation ([Lindblom, 1963](#); [Moon and](#)

---

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: fox.2@osu.edu

Lindblom, 1994), the amount of formant undershoot depends on the interaction between the phonetic context, vowel duration, and the spectral roc. The higher spectral roc is related to faster articulatory movements typically invoked to reach the formant target. An application of this model in concatenative synthesis showed that a better control of spectral roc improves the naturalness and intelligibility of synthesized utterances (Wouters and Macon, 2002b).

Dynamic variations at the vowel target do exist in naturally produced nominal monophthongs and have been found in selective acoustic studies of American, Canadian, and Australian English vowels (e.g., Andruski and Nearey, 1992; Hillenbrand *et al.*, 1995; Watson and Harrington, 1999). Yet, there is no acoustic evidence that vowel-inherent spectral change may actually vary systematically across geographic regions of the country and that the use of time-varying features may be a subject to regional variation. Sociolinguistic work on phonetics and vowel shifts, culminating in the *Atlas of North American English* (Labov *et al.*, 2006), has been primarily concerned with relative positions of vowels in order to create regional maps. While documenting vowel fronting, backing, lowering, raising, centralization, or mergers, formant measurements have been taken typically at one temporal location at the vowel target, referred to as the vowel nucleus. Although this procedure accounts for the regional differences in the positions of vowel nuclei, it does not address the issue of how vowels differ in the extent and nature of dynamic information which may also contribute to regional variation in American English. Crucially, information about how formant frequency changes in time is missing.

There is evidence that the duration of American English vowels varies significantly across regions in the United States (e.g., Clopper *et al.*, 2005; Jacewicz *et al.*, 2007) and so does speech tempo (e.g., Byrd, 1994; Jacewicz *et al.*, 2009). It can be expected that these temporal factors may have a profound effect on formant frequency change in the course of the vowel. As shown by Lindblom and colleagues, there is a complex interaction between vowel duration, consonantal context, and speaking style on formant frequency shifts so that both the position of the vowel in the acoustic space and its spectral dynamics will vary in predictable ways (e.g., Lindblom *et al.*, 2007; Moon and Lindblom, 1994). Expansion or reduction in the vowel space and degree of coarticulation with surrounding consonants are the most observable effects. However, in addition to these and other sources of phonetic variation such as emphatic stress and tempo, regional variation introduces yet another variable to be accounted for in characterizing the acoustic structure of vowels. Cross-dialectal differences in vowel duration, speech tempo, and the extent of formant change within the vowel pose a question of the importance of time-varying information in the differentiation of regional variants. As our understanding progresses, we can ask further questions such as to what extent do speakers of a specific dialect rely on the dynamic aspects of the acoustic signal in identifying vowels as coming from their own dialect.

The aim of the present study was to define the nature of the dynamic spectral change at the targets, that is, the central sections of selected vowels in three distinct regional varieties

of American English spoken in the South (Western North Carolina), in the central region around Columbus, OH, and in the North (Southern Wisconsin). The vowels selected included a true diphthong (/aɪ/), a diphthongized long vowel (/eɪ/), and three lax vowels which exhibited differences in the degree of their diphthongal and temporal properties (/ɪ, ɛ, æ/). The dynamic variations in the formants F1 and F2 were assessed in a set of acoustic measures: vector length (VL), trajectory length (TL), and the spectral roc. Two sources of phonetic variation were included which are known to affect systematically vowel duration: emphatic stress and consonantal context. The study sought to determine the extent to which the dialectal differences in spectral features are present in combination with expected spectral variation as a function of emphasis and context (Lindblom *et al.*, 2007).

## II. METHODS

### A. Speakers

Forty eight women aged 51–65 years participated in the study. They were born, raised, and spent most of their lives in one of three regions in the United States: 16 were from Western North Carolina (the Jackson County area), 16 were from Central Ohio (Columbus area), and 16 were from Southern Wisconsin (Madison area). Defined geographically, these participants created highly homogeneous samples of regional speech, deeply rooted in the regional dialect. According to the *Atlas of North American English* (Labov *et al.*, 2006), these dialects represent Inland South, the Midland, and Inland North, respectively. The recordings were completed in the years 2006–2008. None of the speakers reported any speech disorders. All participants were paid for their efforts.

### B. Speech materials

Five American English vowels were selected for the study: /t, ɛ, e, æ, aɪ/, which varied in their pattern of spectral change (or degree of diphthongization). Each vowel was contained in a target word of the structure /bVts/ and /bVdz/, which yielded the following words: *bits, bets, baits, bats, bites* and *bids, beds, bades, bads, bides*. The target word occurred in a sentence constructed to elicit two levels of vowel emphasis (emphatic and nonemphatic). In these sentences, only the main sentence stress varied, and the position of the target word and its immediate phonetic context remained unchanged. Thus, the proximity of the target word to the changing main sentence stress created the difference in the emphasis of the target word, as exemplified below:

#### *Emphatic*

Sue thinks the small CUTS are deep. No! Sue thinks the small **BITES** are deep.

#### *Nonemphatic*

Sue thinks the small **bites** are WIDE. No! Sue thinks the small **bites** are DEEP.

The use of sentence pairs rather than single sentences ensured fluency in reading, which was essential to examine the variation in the amount of the spectral change in vowels. It has been noted in preliminary studies that some speakers achieved the desired fluency while reading the second sen-

tence. Their first sentence tended to contain hesitations and pauses, which introduced noise to the clarity of exposition of the levels of emphasis. For this reason, only the second sentence in the pair was selected for the present analysis for a total of 1920 sentences (5 vowels  $\times$  2 consonantal contexts  $\times$  2 emphatic positions  $\times$  2 repetitions  $\times$  48 speakers).

### C. Procedure

The testing took place at the university facilities in three locations (North Carolina, Ohio, and Wisconsin). A head-mounted Shure SM10A dynamic microphone was used, positioned at a distance of about 1.5 in. from the speaker's lips. The speaker was seated and was facing a computer monitor. Recordings were controlled by a custom program in MATLAB, which displayed the sentence pair to be read by the speaker and a set of control buttons for the experimenter. The sentences were presented in random order, and the recordings took place in one testing session. Speech samples were recorded and digitized at a 44.1-kHz sampling rate directly onto a hard disk drive. The speaker read the sentence pair placing the main sentence stress on the word in all caps. There was a short practice set completed before the start of the experiment. After recording each sentence pair, the experimenter either accepted and saved the utterance (which occurred most of the time) or re-recorded it in the case of any mispronunciations, disfluencies, or inaccurate stress placement. If the latter took place, the speaker was asked to repeat the utterance as many times as needed.

### D. Acoustic measurements

The set of measurements included vowel duration and the frequencies of F1 and F2 over the course of vowel's duration, which were used to derive further measures of formant movement: VL, TL, and spectral roc. Prior to acoustic analysis, the tokens were digitally filtered and downsampled to 11.025 kHz.

#### 1. Formant frequencies

Measurements of vowel duration served as input for subsequent automated measurements of formant frequencies at five equidistant temporal locations corresponding to the 20%–35%–50%–65%–80% point in the vowel. This was done to eliminate the immediate effects of surrounding consonants on vowel transitions and examine the variation in formant movement spanning over the vowel target. While proportional sampling of formants at two locations close to vowel onset and offset (i.e., 20%–80% or 20%–70%) or three locations including the temporal vowel midpoint (the 50% point) has been used more commonly in several acoustic studies (e.g., Ferguson and Kewley-Port, 2002; Hillenbrand *et al.*, 1995; Hillenbrand *et al.*, 2001), a denser multiple sampling at 4 (Fox, 1983), 9 (Adank *et al.*, 2004), or 16 equidistant points (Van Son and Pols, 1992) has also been done to estimate vowel inherent spectral change. The present use of five equidistant temporal points seeks to characterize the spectral change independent of vowel duration and provide enough information about formant trajectory changes

which may be dialect-specific and may remain unnoticed while sampling the formants at only two or three points.

The frequency change in F1 and F2 over time was measured by centering a 25-ms Hanning window at each temporal location. F1 and F2 values were based on 14-pole linear predictive coding (LPC) analysis and were extracted automatically using a MATLAB program which displayed these values along with the fast fourier transform (FFT) and LPC spectrum and a wideband spectrogram of the entire vowel. In some cases, the formant values were verified using smoothed FFT spectra and wideband spectrograms with formant tracks displayed (using the program TF32, Milenkovic, 2003). Errors in formant estimation in LPC analysis were then hand-corrected.

#### 2. Vowel duration

Standard measures of vowel duration were used (Peterson and Lehiste, 1960; Hillenbrand *et al.*, 1995). Vowel onsets and offsets were located by hand, primarily on the basis of a waveform display with segmentation decisions checked against a spectrogram. Vowel onset was measured from onset of periodicity (at a zero crossing) following the release burst of the stop (if present). In cases where closure remained voiced throughout and there was no evidence of an audible burst release, vowel onset was located at the point which indicated higher amplitude and higher frequency components. Vowel offset for words ending in a voiceless /ts/ was located at the point at which the amplitude of the vowel dropped to near zero (which was also coincident with elimination of all periodicity in the waveform). The vowel offset for words ending in a voiced /dz/ was defined as that point when the amplitude dropped significantly (to near zero). Since any voicing produced during the closure of a voiced stop will have relatively little high frequency energy (Pickett, 1999), this lack of high frequency components will produce a waveform that is relatively sinusoidal showing only slow variations (Olive *et al.*, 1993). When examining the waveforms, both cues were used to identify the location of the stop closure for /d/. All segmentation decisions were later checked and corrected (and then re-checked by a second experimenter) using a custom MATLAB program which displayed the segmentation marks superimposed over a display of the waveform (in two different views: a view that included the entire token and an expanded view that concentrated on the vowel portion only).

#### 3. VL

VL, the length of a vector in F1 by F2 plane, is an indication of the amount of formant change in the course of vowel's duration, typically measured between the 20% and 80% points (Ferguson and Kewley-Port, 2002; Hillenbrand *et al.*, 1995). The assumption is that the longer the vector, the greater the magnitude of formant movement. Diphthongal or diphthongized vowels will have longer vectors than will monophthongs, which corresponds to their greater amount of frequency change. VL is included in the present study to assess its effectiveness as a measure of formant dynamics particularly for vowels in which the direction of formant



movement changes over time. VL is defined as a Euclidean distance (in hertz) between the 20% and 80% temporal points in the vowel in the F1 by F2 plane and is calculated as

$$VL = \sqrt{(F1_1 - F1_5)^2 + (F2_1 - F2_5)^2}. \quad (1)$$

#### 4. TL

Formant TL represents a measure of formant movement which tracks more closely formant frequency change over the course of vowel's duration than the magnitude of formant movement (VL). TL is potentially advantageous to measure the amount of frequency change for diphthongized and vowels whose curved formant tracks resemble a "U-turn" so that the values in the later portion of the vowel return to the values at the vowel's onset. Sampling formant frequencies at five equidistant locations allowed us to calculate TL for each of four separate vowel sections, i.e., 20%–35%, 35%–50%, 50%–65%, and 65%–80%, where the length of one vowel section (VSL) is

$$VSL_n = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2}. \quad (2)$$

The overall formant TL was then defined as a sum of trajectories of four vowel sections:

$$TL = \sum_{n=1}^4 VSL_n. \quad (3)$$

#### 5. Spectral roc

Although TL measure can incorporate the curves in the formant tracks providing a detailed account of formant change, it fails to characterize the amount of frequency change over time. Yet, differences in vowel dynamics are manifested in the way the spectral change varies across vowel's duration. To address this, we first calculated the spectral roc (TL\_roc) over the 60% portion of the vowel which was defined as

$$TL\_roc = \frac{TL}{0.60 \times v\_dur}. \quad (4)$$

In addition, vowel section roc (VSL\_roc) was calculated for each individual vowel section [determined by the temporal location of the five measurement points (20%–35%, 35%–50%, 50%–65%, and 65%–80%)] to compare regions of specific vowels and characterize the nature of the change within a particular region:

$$VSL\_roc_n = \frac{VSL_n}{0.15 \times v\_dur}. \quad (5)$$

It is expected that VSL\_roc will vary not only from section to section within a particular vowel but will also reveal potential differences in the way dialects utilize vowel dynamics for the same vowel "category."

TABLE I. Mean durations of individual vowels (in ms) (s.d.) in emphatic position preceding voiceless (b\_vl) and voiced (b\_vd) consonants.

Vowel	North	North	Ohio	Ohio	Wisconsin	Wisconsin
	Carolina	Carolina				
	b_vl	b_vd	b_vl	b_vd	b_vl	b_vd
/ɪ/	170 (46)	226 (51)	125 (34)	185 (58)	106 (23)	150 (28)
/ɛ/	197 (45)	254 (57)	153 (38)	216 (60)	137 (23)	181 (32)
/e/	210 (49)	268 (57)	183 (36)	263 (63)	174 (29)	252 (44)
/æ/	251 (52)	292 (59)	229 (46)	300 (65)	215 (35)	277 (50)
/aɪ/	239 (39)	295 (52)	197 (38)	291 (68)	175 (26)	274 (52)
Total	214 (55)	267 (61)	178 (52)	251 (77)	162 (46)	227 (67)

### III. RESULTS

#### A. Vowel duration

We begin with the presentation of the results for vowel duration. As displayed in Tables I and II, there were systematic differences in duration as a function of vowel quality, consonantal context, and degree of emphasis. Duration increased progressively with vowel openness which is a well-known intrinsic property of vowels. As also expected, vowels preceding voiced consonants were longer than before voiceless and emphatic vowels were longer than nonemphatic vowels. Of particular interest are differences in vowel duration as a function of dialect. North Carolina speakers produced the longest vowels, followed by Ohio and Wisconsin, respectively.

An analysis of variance (ANOVA) with the within-subject factor vowel, consonantal context and emphasis, and the between-subject factor dialect was used to assess these differences. For all reported significant main effects and interactions, the degrees of freedom for the F-tests were Greenhouse–Geisser adjusted in those cases in which there were significant violations of sphericity. In addition to the significance values, a measure of the effect size—partial eta squared ( $\eta^2$ )—is also reported.

All three within-subject effects were significant and their effect size was strong. As expected, the significant main effect of vowel ( $[F(4, 180) = 674.37, p < 0.001, \eta^2 = 0.937]$ ) reflected the intrinsic differences in the durations of the vowels examined here. The significant effect of consonantal context ( $[F(1, 45) = 326.6, p < 0.001, \eta^2 = 0.879]$ ) confirmed once again that vowel preceding a voiced consonant is longer than vowel preceding a voiceless consonant (means = 213 and 166 ms, respectively). The significant effect of emphasis was

TABLE II. Mean durations of individual vowels (in ms) (s.d.) in nonemphatic position preceding voiceless (b\_vl) and voiced (b\_vd) consonants.

Vowel	North	North	Ohio	Ohio	Wisconsin	Wisconsin
	Carolina	Carolina				
	b_vl	b_vd	b_vl	b_vd	b_vl	b_vd
/ɪ/	135 (33)	158 (41)	91 (24)	127 (40)	88 (18)	114 (37)
/ɛ/	153 (40)	166 (41)	117 (25)	130 (39)	113 (25)	121 (30)
/e/	178 (34)	206 (49)	148 (35)	185 (50)	144 (30)	175 (35)
/æ/	179 (38)	227 (56)	165 (36)	201 (46)	158 (26)	200 (36)
/aɪ/	194 (38)	225 (46)	164 (25)	210 (48)	156 (25)	206 (46)
Total	168 (42)	196 (55)	137 (41)	171 (56)	132 (37)	163 (53)

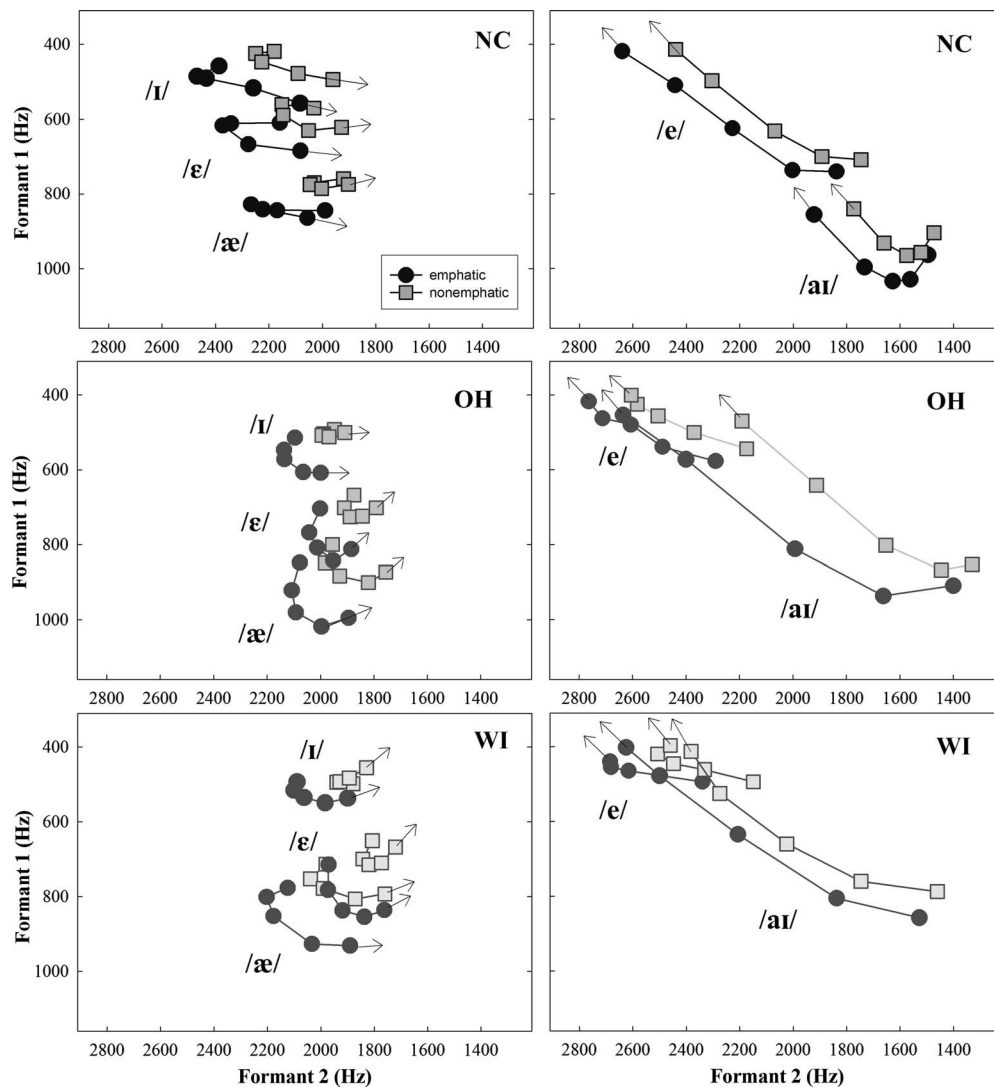


FIG. 1. Mean relative positions of monophthongal (left) and diphthongal (right) vowels and their formant movement measured at five equidistant points in the central 60% portion of the vowel. Shown are emphatic and nonemphatic vowels produced in the bVts context (“voiceless”) by female speakers of three dialectal varieties of American English (spoken in Western North Carolina, Central Ohio, and Southern Wisconsin).

manifested in longer durations of emphatic vowels as compared to nonemphatic ( $[F(1,45)=177.26, p<0.001, \eta^2=0.798]$ , means=217 and 161 ms, respectively). Interestingly, mean differences in vowel duration as a function of either emphasis or consonantal context were comparable (56 and 47 ms, respectively), indicating that consonantal context effects on vowel duration can be as great as the effects of emphasis.

The main effect of dialect was significant ( $[F(2,45)=6.06, p=0.005, \eta^2=0.213]$ ) although its effect size was smaller than that for the within-subject factors. Subsequent *post-hoc* analyses using separate ANOVAs which included two dialects only showed that Wisconsin and North Carolina vowels differed significantly from one another (means=171 and 211 ms, respectively). However, Ohio vowels (means=185 ms) did not differ significantly from either Wisconsin or North Carolina vowels. These cross-dialectal differences in vowel duration are consistent with the results reported in [Jacewicz et al. \(2007\)](#) for young adults, confirming that dialectal differences in vowel duration do exist (at least for selected regions) and are independent of speaker age.

Several interactions were significant although their nature and small effect size do not warrant a separate discussion. One significant interaction between context and emphasis deserves mention given its large effect size  $[F(1,45)=152.29, p<0.001, \eta^2=0.772]$ . The interaction arose from the fact that emphatic vowels in the context of voiced consonants were substantially longer (72 ms or 41%) than nonemphatic vowels in this environment whereas the emphasis-related difference for vowels preceding voiceless consonants was smaller (39 ms or 27%).

## B. Formant movement

Turning to formant analysis, Figs. 1 and 2 display relative positions in the  $F1 \times F2$  plane and formant movement of vowels preceding voiceless and voiced consonants, respectively. The left panels show “monophthongal” vowels /i, ε, æ/ and the right panels the “diphthongal” /e, aɪ/. Direction of formant movement is indicated by arrows.

Based on visual inspection, there is a substantial variation in formant dynamics across individual vowels and dia-

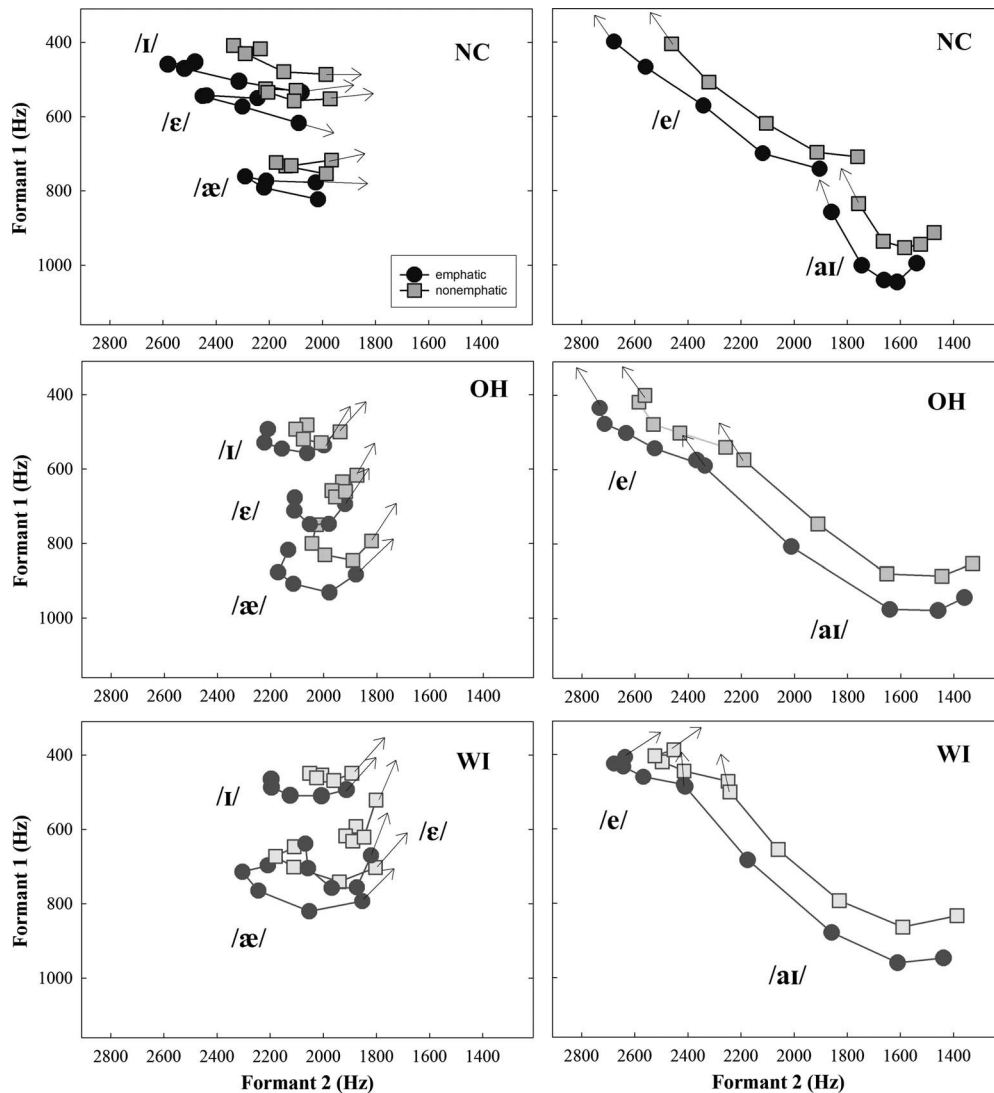


FIG. 2. Mean values of the vowels in the bVdz (“voiced”) context (see Fig. 1 legend for details).

lects. North Carolina /i, ε, æ/ are the most fronted with the nature of their formant movement distinct from both Ohio and Wisconsin vowels. Across all dialects, emphatic vowels are more peripheral and show more formant movement than nonemphatic vowels. Cross-dialectal differences are particularly evident for /e, aɪ/. The North Carolina /e/ is the most diphthongal and the Wisconsin /e/ may even be regarded as a monophthong given its small amount of change in F1. The diphthong /aɪ/, on the other hand, is relatively monophthongal in North Carolina but shows a great amount of spectral change in both Ohio and Wisconsin. The Wisconsin /æ/ is raised due to the Northern Cities Shift and it can be seen that its nonemphatic variant has considerable overlap with the emphatic /ε/.<sup>1</sup>

It needs to be underscored that the formant track dynamics displayed in Figs. 1 and 2 is plotted from measurements at five temporally equidistant points during a vowel. Thus, these frequency measurements are time-normalized across all vowels and do not reflect differences in vowel duration. This issue will be addressed subsequently.

### C. VL

The first measure applied to assess the present variation in formant dynamics is VL (e.g., Ferguson and Kewley-Port, 2002; Hillenbrand *et al.*, 1995; Hillenbrand and Nearey, 1999). As Fig. 3 shows, VLs are smaller for some vowels such as /i, ε/ but every vowel exhibits at least some amount of spectral change. There are clear VL differences as a function of dialect, especially between North Carolina vowels and those from the two Midwestern dialects. A separate repeated-measures ANOVA was conducted for each vowel<sup>2</sup> with the within-subject factors consonantal context and emphasis. Dialect was included as the between-subject factor. In general, all three main effects were significant. One exception was the vowel /ε/, whose VLs did not differ significantly as a function of dialect. Table III summarizes the results of the analyses. As can be seen, the effect size was typically greater for the main effect of emphasis compared to the main effect of consonantal context. For all vowels, VLs were significantly longer for emphatic vowels than for nonemphatic. The context effects were more variable, indicating longer

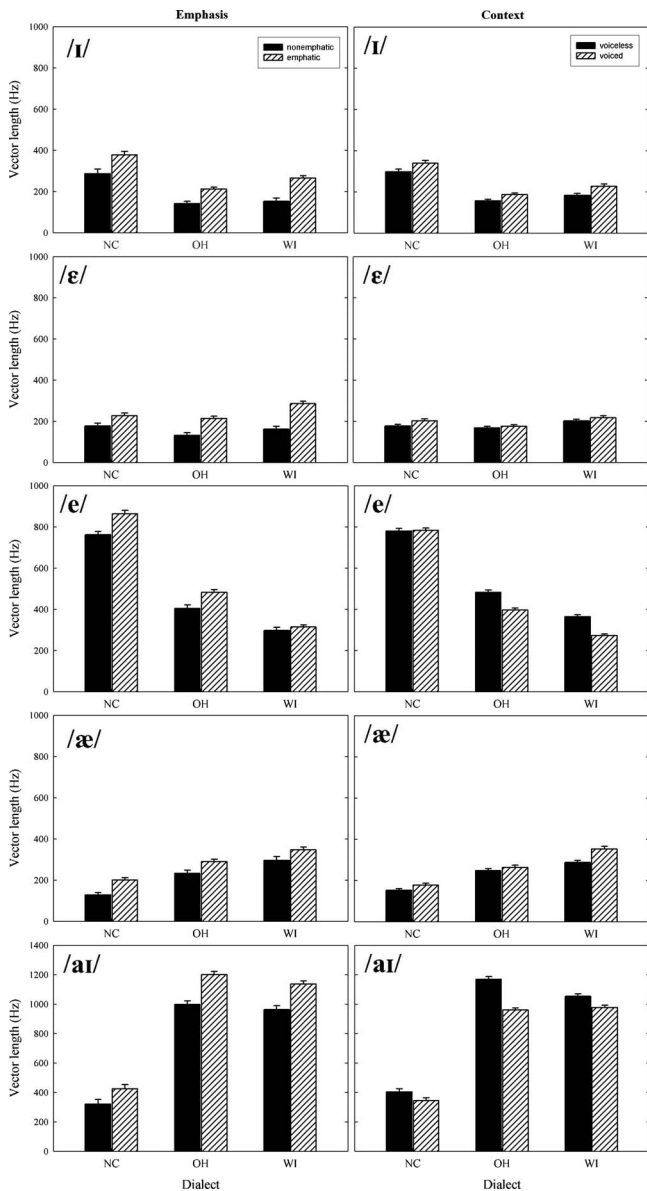


FIG. 3. Mean values (s.e.) for VL, i.e., F1 and F2 frequencies change between the 20%–80% temporal point, for each vowel in each dialect as a function of vowel emphasis and consonantal context.

VLs when the vowel was followed by voiced consonants in the case of /ɪ, ɛ, æ/ and longer VLs when it was followed by voiceless consonants for /e, aɪ/. The effects of dialect were particularly strong for the vowels /e, aɪ/ due to the fact that North Carolina VLs were clearly different from both Midwestern variants. The North Carolina /e/ had the longest VL which was more than twice that of the Wisconsin variant, with Ohio vowel falling in between. For the diphthong /aɪ/, Ohio variant had the longest VL, about three times that of North Carolina /aɪ/ which is well known for being relatively monophthongal in this regional variety of English.

It can be expected that the large differences between the North Carolina vowels and the vowels from the two Midwestern varieties will produce a significant main effect of dialect. However, the differences between the Ohio and Wisconsin vowels themselves may be too small to reach significance. Additional repeated-measures ANOVAs were used to examine the significance of all three factors (emphasis, context, and dialect) for Ohio and Wisconsin vowels while excluding North Carolina from the analyses. As expected, the results showed significant main effects of emphasis and context for each of the five vowels. However, the main effect of dialect was significant only for /e/ [ $F(1, 30)=19.55$ ,  $p < 0.001$ ,  $\eta^2=0.394$ ], indicating longer VL for Ohio variant compared to Wisconsin.

In summary, the present results show that VLs varied significantly with vowel emphasis and consonantal context, and dialectal differences were also apparent, at least between North Carolina and Midwestern vowels. However, one issue needs to be resolved before accepting VL as a measure which characterizes the true amount of frequency change in the course of vowel's duration. In particular, one can argue that VL, in fact, underestimates the amount of spectral change in a vowel and may lead to false interpretations of the nature of the spectral change being examined. To exemplify the point, we will now consider two examples of North Carolina vowels: /æ/ and /e/.

The left panel of Fig. 4 shows the North Carolina variant of /æ/ redrawn here from Fig. 1 for the purposes of illustration. VL is a measure of formant frequency change between the 20% and 80% points in the vowel. As evident, VL fails to

TABLE III. Summary of significant main effects and interactions from repeated measures ANOVAs for VL. Shown are partial eta squared values ( $\eta^2$ ).—not significant, vd=voiced, vl=voiceless, e=emphatic, ne=nonemphatic, NC=North Carolina, OH=Ohio, and WI=Wisconsin.

	/ɪ/	/ɛ/	/e/	/æ/	/aɪ/
Context	0.269 <sup>a</sup>	0.152 <sup>b</sup>	0.515 <sup>a</sup>	0.284 <sup>a</sup>	0.267 <sup>a</sup>
	vd > vl	vd > vl	vl > vd	vd > vl	vl > vd
Emphasis	0.474 <sup>a</sup>	0.517 <sup>a</sup>	0.375 <sup>a</sup>	0.325 <sup>a</sup>	0.510 <sup>a</sup>
	e > ne	e > ne	e > ne	e > ne	e > ne
Dialect	0.305 <sup>a</sup>	—	0.818 <sup>a</sup>	0.292 <sup>a</sup>	0.800 <sup>a</sup>
	NC > WI > OH	—	NC > OH > WI	WI > OH > NC	OH > WI > NC
Context × Emphasis	0.092 <sup>c</sup>	—	—	—	0.111 <sup>c</sup>
Context × Dialect	—	—	0.341 <sup>a</sup>	—	—
Emphasis × Dialect	—	—	0.155 <sup>c</sup>	—	—
Con × Emp × Dialect	—	—	0.136 <sup>c</sup>	—	—

<sup>a</sup>  $p < 0.001$ .

<sup>b</sup>  $p < 0.010$ .

<sup>c</sup>  $p < 0.050$ .



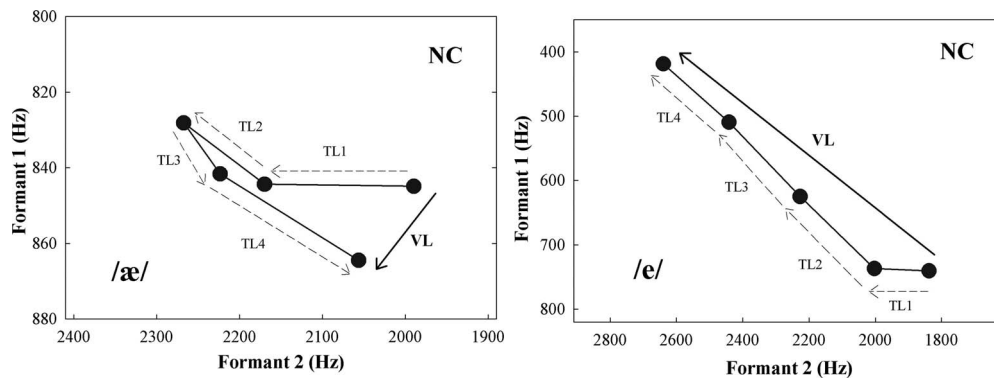


FIG. 4. Measurement of VL and total TL for North Carolina variant of /æ/ (left) and /e/ (right) in emphatic positions redrawn from Fig. 1.

account for the actual formant movement over time. The length of the entire formant trajectory consists here of four sections (TL1–TL4), each corresponding to formant change between two consecutive measurement points (20%–35%, 35%–50%, 50%–65%, and 65%–80%). Because the trajectory of North Carolina /æ/ is U-shaped (which reflects the “Southern drawl”), the VL estimate is particularly inadequate to measure this type of spectral change. Yet, VL can be quite accurate in assessing diphthongal changes such as for North Carolina /e/ shown in the right panel. This vowel, also redrawn from Fig. 1, shows an almost linear spectral change across its four sections. Thus, the estimated total trajectory change can be approximated relatively well by the length of the vector which expresses a linear distance between the 20%- and 80% points.

In summary, VL does appear to capture some aspects of formant movement and is rather reliable as a measure of linear trajectory change. However, formant trajectory shapes can vary cross-dialectally in ways impossible to characterize by the use of VL. The North Carolina /æ/ is the most fitting example. It seems that computing the length of the entire trajectory, i.e., approximated by the multiple-point sampling, may account more reliably for the extent of spectral change in a vowel. Section IV will address this possibility.

#### D. TL

The total TL, consisting of the sum of TLs of the four vowel sections, is expected to provide a more detailed estimate of formant change. Figure 5 shows mean TL values for each vowel broken down by emphasis and consonantal context. As expected, the TL values are greater than those for VL in Fig. 3. As it was done for the VL measure, a repeated-measures ANOVA with the within-subject factors emphasis and consonantal context and between-subject factor dialect was conducted for each vowel.

The main effects of emphasis and consonantal contexts were significant, and the general results were in accord with those for VL: emphatic vowels had significantly greater TLs than nonemphatic vowels, the vowels /i, ε, æ/ had longer TLs when followed by voiced consonants, and /e, ai/ had longer TLs when followed by voiceless consonant. Table IV summarizes the results of the analyses.

The effects of dialect were somewhat different for TLs, however. Although the main effect of dialect was significant for the vowels /i, e, ai/ and the order of dialectal variants in

terms of the amount of the spectral change were in agreement with the results for VL (including the significant difference between the Ohio and Wisconsin /e/), discrepancies between the two measures were found for the vowels /ε/ and

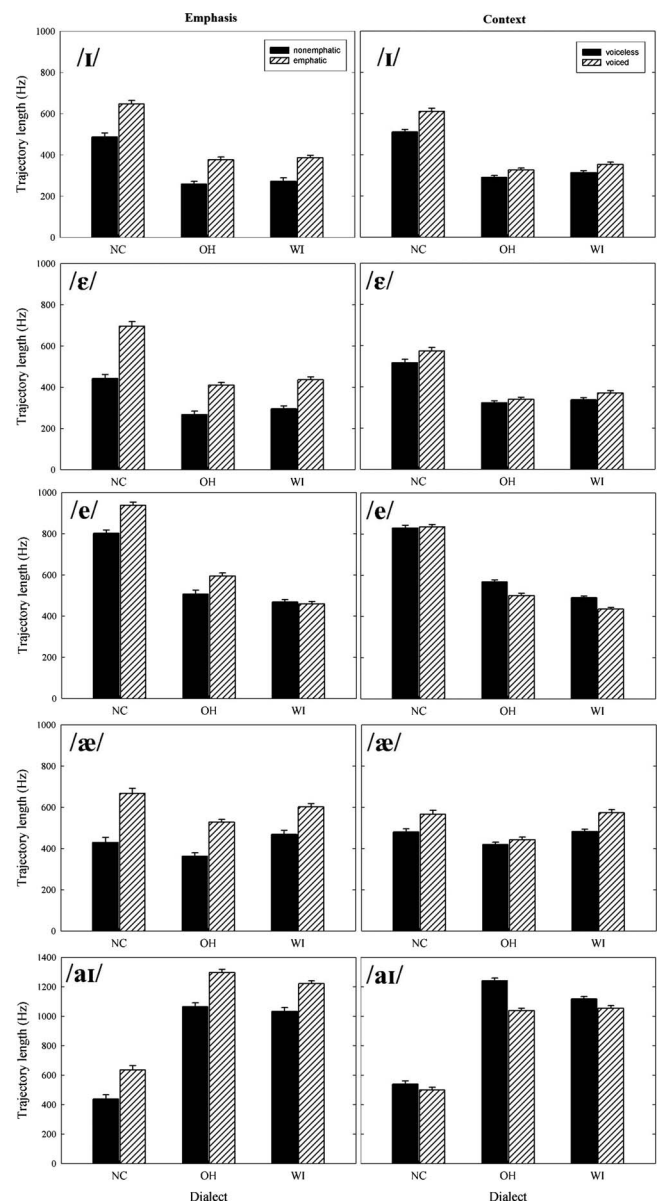


FIG. 5. TL, i.e., sum of the VSLs of four vowel sections over the central 60% of the vowel. Shown are mean values (s.e.) for each vowel in each dialect as a function of vowel emphasis and consonantal context.



TABLE IV. Summary of significant main effects and interactions from repeated measures ANOVAs for total TL. Shown are partial eta squared values ( $\eta^2$ ).—not significant, vd=voiced, vl=voiceless, e=emphatic, ne=nonemphatic, NC=North Carolina, OH=Ohio, and WI=Wisconsin.

	/i/	/ɛ/	/e/	/æ/	/aɪ/
Emphasis	0.581 <sup>a</sup>	0.651 <sup>a</sup>	0.319 <sup>a</sup>	0.680 <sup>a</sup>	0.623 <sup>a</sup>
	e > ne	e > ne	e > ne	e > ne	e > ne
Context	0.503 <sup>a</sup>	0.118 <sup>b</sup>	0.266 <sup>a</sup>	0.402 <sup>a</sup>	0.280 <sup>a</sup>
	vd > vl	vd > vl	vl > vd	vd > vl	vl > vd
Dialect	0.670 <sup>a</sup>	0.516 <sup>a</sup>	0.804 <sup>a</sup>	—	0.737 <sup>a</sup>
	NC > WI > OH	NC > WI > OH	NC > OH > WI	—	OH > WI > NC
Context × Emphasis	—	—	—	—	—
Context × Dialect	0.138 <sup>b</sup>	—	0.180	—	0.180 <sup>b</sup>
Emphasis × Dialect	—	0.144 <sup>b</sup>	0.257 <sup>c</sup>	—	—
Cont × Emp × Dialect	0.139 <sup>b</sup>	—	—	—	—

<sup>a</sup> $p < 0.001$ .

<sup>b</sup> $p < 0.050$ .

<sup>c</sup> $p < 0.010$ .

/æ/. In particular, there was no significant effect of dialect for the VL measure for /ɛ/ [ $F(2, 45) = 1.56, p = 0.221, \eta^2 = 0.065$ ], whereas dialect was significant for TL [ $F(2, 45) = 23.95, p < 0.001, \eta^2 = 0.516$ ], showing greater TLs for North Carolina /ɛ/ (mean = 571 Hz) than for either Wisconsin (mean = 366 Hz) or Ohio (mean = 341 Hz). For the vowel /æ/, the pattern was reversed: the main effect of dialect was significant for the VL measure [ $F(2, 45) = 9.29, p < 0.001, \eta^2 = 0.292$ ], showing greatest VLs for Wisconsin (mean = 317 Hz) followed by Ohio and North Carolina (means = 259 and 164 Hz, respectively). For the TL measure, dialect was not significant [ $F(2, 45) = 3.15, p = 0.053, \eta^2 = 0.123$ ] and North Carolina /æ/ had slightly greater TL (mean = 549 Hz) than Wisconsin (mean = 535 Hz), with Ohio falling last (444 Hz). Clearly, these discrepancies arose from underestimating the amount of formant change by the VL measure due to the change in the direction of formant curves.

To compare the results of the two measures, i.e., VL and TL, separate repeated-measures ANOVAs were used for each vowel and for each dialect with the within-subject factors formant change (VL, TL), emphasis, and consonantal context. Table V summarizes the results for the main effect of formant change.

As can be seen, the differences between VL and TL were highly significant for each vowel in each dialect. Next to the effect size, the table lists in parentheses the percentage of underestimation of formant change by the VL measure. The underestimation was found to be as great as 70% for the North Carolina /æ/ and as small as 7%–8% for the Ohio and Wisconsin /aɪ/ and North Carolina /e/. For the remaining

vowels, the VL underestimation ranged from 19% to 65%. These results show an advantage of the TL measure over VL, especially for vowels which exhibit a change in the direction of formant movement. The general picture of TL advantage for each vowel averaged across emphasis levels and consonantal contexts can be found in Fig. 6. For each dialect, the VL underestimation of formant change for the vowels /i, ɛ, æ/ is considerably greater than for the diphthongal vowels /e, aɪ/. These differences were found for each dialect, indicating that the TL measure reflects dialect-specific spectral change in vowels quite well.

In summary, the statistical evidence along with the graphic displays suggests that VL does not account reliably for the dialectal differences. The extent of formant movement is better characterized by a TL measure, which utilizes formant measurements sampled at multiple points in a vowel.

## E. Spectral roc

Although the TL measure appears to be more reliable in addressing dialectal differences, the measurement points are time normalized and indicate only relative positions across the vowel. This, of course, fails to account for how quickly (or slowly) these formant frequency changes occur in time. Yet, there may be important dynamic differences across dialects, contexts, and speaker age that relate to such spectro-temporal changes. The spectral roc measure presented here will allow us to make these comparisons.

TABLE V. Summary of the significant main effect of formant change (VL vs TL) from repeated measures ANOVAs. Shown are partial eta squared values ( $\eta^2$ ). The values in parentheses indicate percentage of underestimation of formant movement by the VL measure.

	/i/	/ɛ/	/e/	/æ/	/aɪ/
North Carolina	0.895 <sup>a</sup> (42)	0.893 <sup>a</sup> (65)	0.792 <sup>a</sup> (7)	0.876 <sup>a</sup> (70)	0.855 <sup>a</sup> (31)
Ohio	0.907 <sup>a</sup> (43)	0.898 <sup>a</sup> (49)	0.744 <sup>a</sup> (19)	0.917 <sup>a</sup> (42)	0.896 <sup>a</sup> (7)
Wisconsin	0.910 <sup>a</sup> (36)	0.916 <sup>a</sup> (39)	0.814 <sup>a</sup> (34)	0.945 <sup>a</sup> (41)	0.878 <sup>a</sup> (8)

<sup>a</sup> $p < 0.001$ .

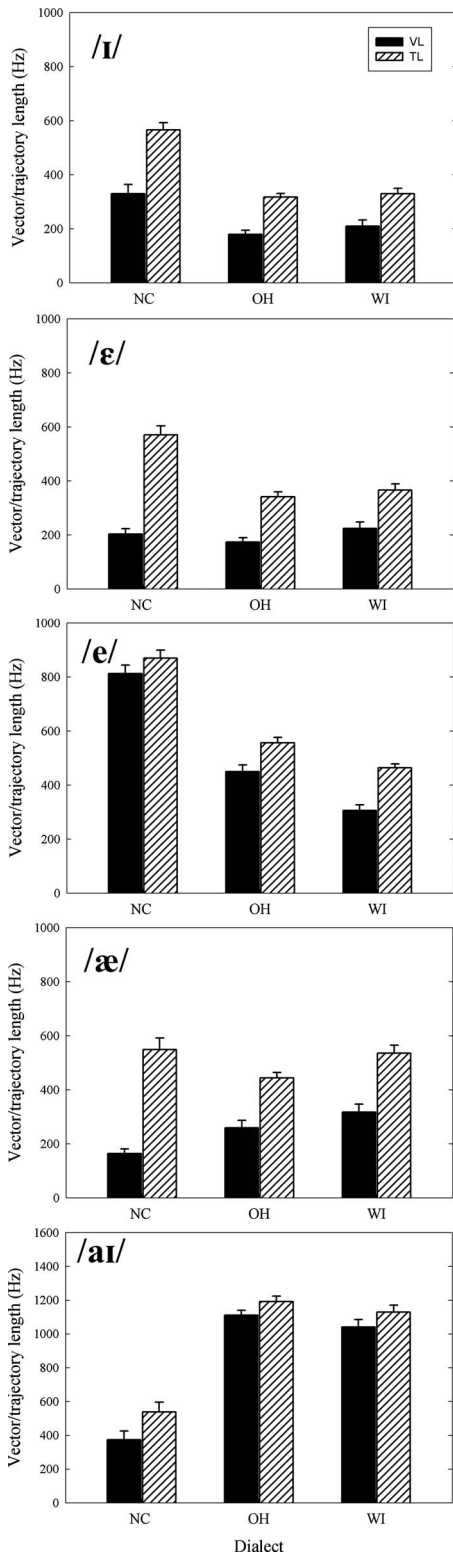


FIG. 6. A comparison of VL and TL for each vowel and each dialect. Shown are mean values (s.e.) averaged across emphasis levels and consonantal context.

Shown in Fig. 7 is the spectral roc for the five vowels in both emphatic and nonemphatic positions in voiced and voiceless contexts for each of the three dialects. As might be expected, overall spectral roc varies as a function of vowel category. The mean values were highest for the diphthong /aɪ/ in both Wisconsin (10.1 Hz/ms) and Ohio (9.8 Hz/ms)

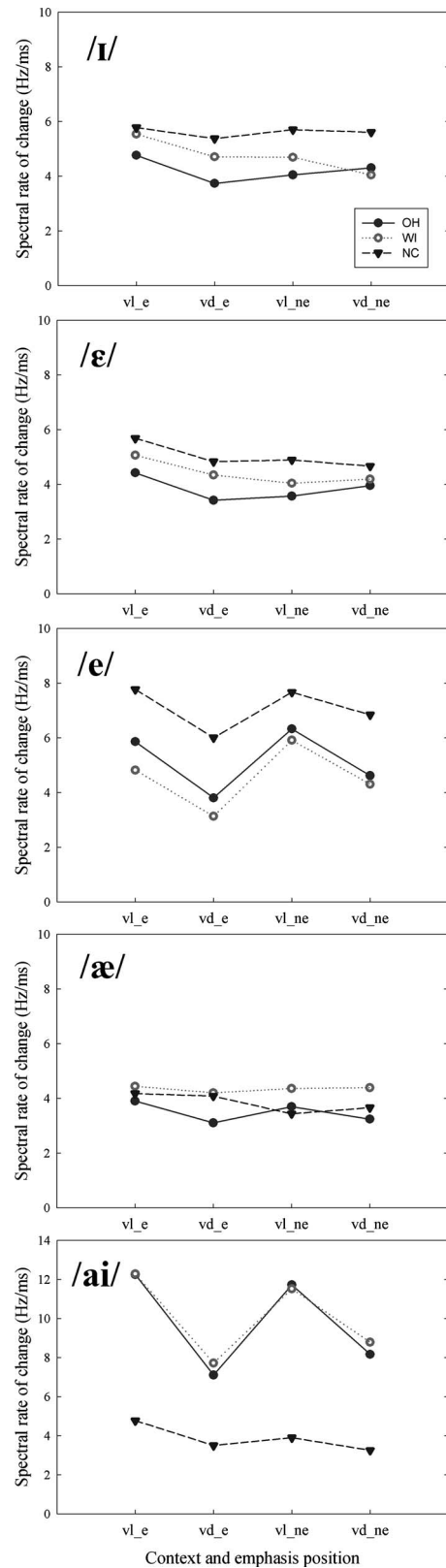


FIG. 7. Mean spectral roc at the targets of vowels in variable emphasis positions (e=emphatic, ne=nonemphatic) and consonantal contexts (vl=voiceless, vd=voiced) across the dialects.

varieties (but not in North Carolina, 3.9 Hz/ms) and lowest for the vowel /æ/ in each of the three dialects (4.4, 3.5, and 3.8 Hz/ms, respectively). Dialectal differences were particularly evident in the case of /e/ which had the highest spectral

TABLE VI. Summary of significant main effects and interactions from repeated measures ANOVAs for spectral roc. Shown are partial eta squared values ( $\eta^2$ ).—not significant, vd=voiced, vl=voiceless, e=emphatic, ne=nonemphatic, NC=North Carolina, OH=Ohio, and WI=Wisconsin.

	/i/	/ɛ/	/e/	/æ/	/aɪ/
Emphasis	—	0.094 <sup>a</sup> e > ne	0.262 <sup>b</sup> ne > e	—	—
Context	0.163 <sup>c</sup> vl > vd	0.160 <sup>c</sup> vl > vd	0.754 <sup>b</sup> vl > vd	0.089 <sup>a</sup> vl > vd	0.783 <sup>b</sup> vl > vd
Dialect	0.286 <sup>c</sup> NC > WI > OH	0.211 <sup>c</sup> NC > WI > OH	0.484 <sup>b</sup> NC > OH > WI	0.154 <sup>a</sup> WI > NC > OH	0.745 <sup>b</sup> WI > OH > NC
Context × Emphasis	0.134 <sup>a</sup>	0.288 <sup>b</sup>	0.106 <sup>a</sup>	0.089 <sup>a</sup>	0.353 <sup>b</sup>
Context × Dialect	—	—	—	0.146 <sup>a</sup>	0.465 <sup>b</sup>
Emphasis × Dialect	—	—	—	—	—
Cont × Emp × Dialect	—	—	—	—	—

<sup>a</sup> $p < 0.05$ .

<sup>b</sup> $p < 0.001$ .

<sup>c</sup> $p < 0.010$ .

roc among all North Carolina vowels (mean=7.1 Hz/ms) and second highest among the Ohio vowels (mean=5.1 Hz/ms). In Wisconsin, however, the mean value was lower (4.5 Hz/ms) and it was comparable with roc for /ɛ/ and /æ/ (both 4.4 Hz/ms).

A separate repeated-measures ANOVA with the within-subject factors emphasis and consonantal context and the between-subject factor dialect was conducted for each vowel. As summarized in Table VI, vowel emphasis did not have a strong effect on spectral roc. Rather, it was the consonantal context that affected the spectral roc of all vowels in a systematic way: without exception, vowels preceding voiceless consonants had higher spectral roc than when preceding voiced consonants. This effect was particularly strong for the vowels /e, aɪ/. The main effect of dialect was significant for each vowel. For the vowels /i, ɛ, e/ North Carolina had the highest spectral roc among the three dialects. For /æ, aɪ/, the spectral roc was highest in the variety of English spoken in Wisconsin. Also significant for each vowel was the interaction between context and emphasis. This interaction, although not particularly strong, was manifested somewhat differently for monophthongal /i, ɛ, æ/ and diphthongal /e, aɪ/ vowels.<sup>3</sup>

Of particular interest to this study are dialectal differences in the spectral roc which persisted when two additional sources of contextual variation in roc were included, i.e., degree of vowel emphasis and the type of consonantal context. Clearly, changes in spectral roc arise from differences in vowel duration, TL or a combination of the two. In an attempt to better understand the contribution of each source of this variation and explain the obtained patterns, we will now examine proportional differences in vowel duration and TL which arose from the two contextual factors, vowel emphasis and consonantal context.

Listed in Table VII are changes in vowel duration and TL as a function of vowel emphasis. Of interest are percentages of reduction in vowel duration in nonemphatic positions and corresponding reduction in TL for each vowel in each dialect. The general tendency is that the proportion of reduction in duration of nonemphatic vowels corresponds roughly to the proportion of reduction in their TLs, which will not

affect their spectral roc. This would explain the lack of significant effect of emphasis on the spectral roc, at least for three out of five vowels. A different outcome was found for the consonantal context effects, as summarized in Table VIII. Here the relative reduction in TL for vowels in voiceless context tends to be smaller than the reduction in vowel duration, which produces higher spectral roc in the voiceless context compared to the voiced context. These results are in accord with the number of significant effects of consonantal context on the spectral roc of individual vowels (compare Table VI).

A word of caution against an exclusive reliance on these general trends in explaining the spectral roc results is needed, however. Several factors interact here, and each has some impact on the movement of articulators which is the underlying source of variation in the spectral roc. In some cases, it is the dialect-specific spectral change that interacts differently with contextual factors. To illustrate the point, we will consider one example here, that of proportional reductions in both vowel duration and TL for the vowel /æ/, which have been found to vary across dialects.

As Table VIII indicates, the North Carolina variant does not increase its spectral roc in the voiceless context. Rather, spectral roc increases in emphatic position which has a greater TL (and smaller decrease in duration) as compared to the nonemphatic position (see Table VII). However, consonantal context (and not emphasis) affected spectral roc of the Ohio /æ/ which increased in the voiceless context due to the small reduction in TL. Finally, the proportional reductions as a function of both emphasis and context did not vary much for the Wisconsin variant, suggesting that, in this dialect, spectral roc of /æ/ does not change across emphatic positions and contexts. Results of separate ANOVAs used for each dialect support this explanation. The main effect of emphasis was significant for the North Carolina /æ/ ( $[F(1, 15)=4.97, p=0.041, \eta^2=0.249]$ ) and indicated higher spectral roc in emphatic position. The main effect of context was significant for Ohio /æ/ ( $[F(1, 15)=7.15, p=0.017, \eta^2=0.323]$ ), showing higher spectral roc in the voiceless context. Finally, nei-

TABLE VII. Changes in mean values of vowel duration (vow dur) and TL as a function of vowel emphasis along with percentages of their reductions in nonemphatic relative to emphatic positions.

Vowel	Vow dur emphatic (ms)	Vow dur nonemphatic (ms)	Vow dur reduction ne < e (%)	TL emphatic (Hz)	TL nonemphatic (Hz)	TL reduction ne < e (%)
North Carolina						
/i/	197.2	146.7	25.6	645.4	486.5	24.6
/ɛ/	225.5	159.4	29.3	699.3	442.0	36.8
/e/	239.0	191.7	19.8	936.8	801.5	14.4
/æ/	271.4	202.7	25.3	667.4	429.6	35.6
/aɪ/	267.1	209.2	21.7	634.7	442.3	30.3
Ohio						
/i/	157.1	109.2	30.5	376.8	257.7	31.6
/ɛ/	186.7	123.9	33.6	411.9	269.7	34.5
/e/	225.6	168.3	25.4	595.4	517.6	13.1
/æ/	266.4	183.6	31.1	521.9	365.8	29.9
/aɪ/	245.1	188.4	23.1	1308.9	1075.0	17.9
Wisconsin						
/i/	128.5	101.1	21.3	386.3	272.4	29.5
/ɛ/	158.6	117.2	26.1	436.6	294.6	32.5
/e/	213.0	159.4	25.2	460.2	468.0	-1.7
/æ/	246.3	178.9	27.4	602.2	468.4	22.2
/aɪ/	224.7	181.0	19.5	1220.5	1036.5	15.1

ther emphasis nor context was significant for the Wisconsin variant. No other effects and interactions were significant in these analyses.

These results support the claim that, as a measure, spectral roc is sensitive to dialectal differences in vowel dynamics and can provide details of complex interactions of several

factors. Since roc does not require extensive computations, it can be used effectively in analyzing a larger corpus.

#### IV. DISCUSSION

The present study sought to characterize the nature of the dynamic spectral change found in the targets of selected

TABLE VIII. Changes in mean values of vowel duration (vow dur) and TL as a function of consonantal context along with percentages of their reductions in voiceless relative to voiced contexts.

Vowel	Vow dur voiced (ms)	Vow dur voiceless (ms)	Vow dur reduction vl < vd (%)	TL voiced (Hz)	TL voiceless (Hz)	TL reduction vl < vd (%)
North Carolina						
/i/	191.4	152.6	20.3	618.6	513.3	17.0
/ɛ/	209.8	175.0	16.6	585.7	555.6	5.1
/e/	236.9	193.9	18.1	871.0	867.3	0.4
/æ/	259.3	214.8	17.1	603.9	493.1	18.3
/aɪ/	259.8	216.6	16.6	524.2	552.8	-5.5
Ohio						
/i/	156.8	109.5	30.2	350.3	284.2	18.9
/ɛ/	173.5	137.1	21.0	354.9	326.7	7.9
/e/	225.1	168.7	25.0	526.5	586.5	-11.4
/æ/	251.4	198.5	21.1	459.1	428.5	6.7
/aɪ/	251.3	182.2	27.5	1086.0	1298.0	-19.5
Wisconsin						
/i/	132.2	97.4	26.3	348.2	310.5	10.8
/ɛ/	150.8	125.0	17.1	387.4	343.7	11.3
/e/	213.2	159.1	25.4	427.9	500.4	-17.0
/æ/	238.8	186.5	21.9	587.7	482.9	17.8
/aɪ/	239.8	165.8	30.8	1092.4	1164.5	-6.6



American English vowels in three distinct dialectal regions in the United States. The results are encouraging in that we are beginning to find ways to better understand vowel dynamics across different American English dialects. In particular, we found cross-dialectal differences in vowel duration, in the extent of spectral change in formant trajectories, and in the spectral roc.

Although our primary goal was to find a set of effective measures which would reveal systematic differences among the regional variants, the study also gained more insights into positional relations within the vowel system of each dialect. As seen in Figs. 1 and 2, Ohio and Wisconsin variants of /ɛ, æ/ tend to spectrally overlap when variable emphasis is taken into consideration: the emphatic /ɛ/ approximates the position of the nonemphatic /æ/. This is not the case for North Carolina, where /ɛ/ and /æ/ are clearly separated under variable emphasis conditions and a possibility of an overlap arises for /ɪ, ɛ/ rather. The nature of formant dynamics is also highly variable across the dialects, and the magnitude of formant movement can vary dramatically such as for the diphthongal vowels /e, ai/.

### A. Characterizing the variation in formant dynamics

We first turned to an established procedure of estimating the magnitude of formant frequency change between the 20% and 80% temporal points in a vowel (VL). Although some of the spectral variation could be accounted for by this measure, we excluded it from further consideration as it did not provide a satisfying characterization of the most dynamic spectral changes in several of the vowels. In particular, the extent of formant movement was greatly underestimated for vowels in which the direction of this movement in the F1 by F2 space changes over time. We found that the total trajectory change (TL) over multiple temporal locations for the vowel center represents more adequately the magnitude of formant movement. Relating the spectral change over the total trajectory to the time necessary to execute this formant movement, we computed the spectral roc which provides another view of the time-varying information in a vowel.

Two phonetic factors that affect vowel duration, emphatic stress and the voicing status of the consonant that follows the vowel, were systematically varied in this study. Entered as within-subject factors, the two sources of variation were found to interact with formant movement in somewhat different way: while both the emphatic vowel and vowel preceding a voiced consonant had longer durations, the spectral roc was significantly higher for the shorter vowel followed by a voiceless consonant and not for the shorter nonemphatic vowel. The small effect size of emphasis found in the analyses of spectral roc will need to be investigated separately in greater detail. We can only speculate that this variation comes from the way emphatic stress is brought about by vowel-specific articulatory actions.<sup>4</sup>

Apart from the variation in vowel dynamics coming from phonetic sources, dialect was found to be a strong source of variation in vowel-inherent spectral change. The effects of dialect were found for each vowel examined in the present study. As an example, an interesting relationship be-

tween vowel duration and the dialect-specific nature of formant trajectory change was found for the vowel /ɪ/. North Carolina /ɪ/ was longer, had a greater TL, and faster spectral roc (means: 172 ms, 566 Hz, 5.6 Hz/ms) than Ohio /ɪ/ which was shorter, had a smaller TL, and slower spectral roc (means: 133 ms, 317 Hz, 4.2 Hz/ms). These differences stem from the nature of the dynamic formant changes in each dialect which, in very general terms, are brought about by faster articulatory gestures in order to produce the North Carolina vowel and comparatively slower gestures in a slightly diphthongal variety of the Ohio /ɪ/.

The spectral roc measure used in this study is just one possible measure which, to some extent, reflects speed of articulatory movement over the course of the vowel's target. Although this measure can only give an indirect indication of the speed of specific articulators underlying the production of diphthongal and quasi-diphthongal changes, it is nevertheless useful in estimating the average pace of formant movement during the central 60% of the total spectral change. A related measure of spectral change, although assessing F2 velocity only, was used in Moon and Lindblom (1994). A more detailed measurement such as by fitting linear regression lines to the formants and computing the slopes (Wouters and Macon, 2002a) will be problematic in this particular set of data because of the changing directionality of the formant movement. While Wouters and Macon (2002a) studied liquid-vowel and vowel-liquid transitions and diphthong transitions in the productions of one speaker, this approach will not be effective in dealing with the type of spectral changes such as those found in North Carolina vowels. The present approach, being relatively easy to implement, can be more readily used in a sociophonetic setting which, by definition, must involve a larger corpus of data. Having established the types of variation in formant trajectories that can be expected in cross-dialectal data in terms of TL, directionality, and curvature, a refinement of the current measures will be undertaken in order to address the changes in the direction of formant movement. In particular, parametrization procedures can be used (e.g., Harrington, 2006; Harrington *et al.*, 2008; Hillenbrand *et al.*, 2001; Morrison, 2009; Zahorian and Jagharghi, 1993) in order to model the various trajectory shapes.

It is clear that the nature and amount of spectral change for vowels studied here can be characterized quite effectively when formant trajectories are sampled at multiple time points rather than at one temporal location at the vowel target. The use of five temporal locations estimates the dynamic trajectory to the extent that time-normalized spectral variation can be assessed rather accurately. The addition of the time dimension and inclusion of the spectral roc provides further insights as to how vowel-inherent spectral change differs for individual vowels across several regional variants. Thus, the combination of the three basic acoustic parameters (TL, vowel duration, and spectral roc) can be effective in characterizing the regional variation in American English vowels.

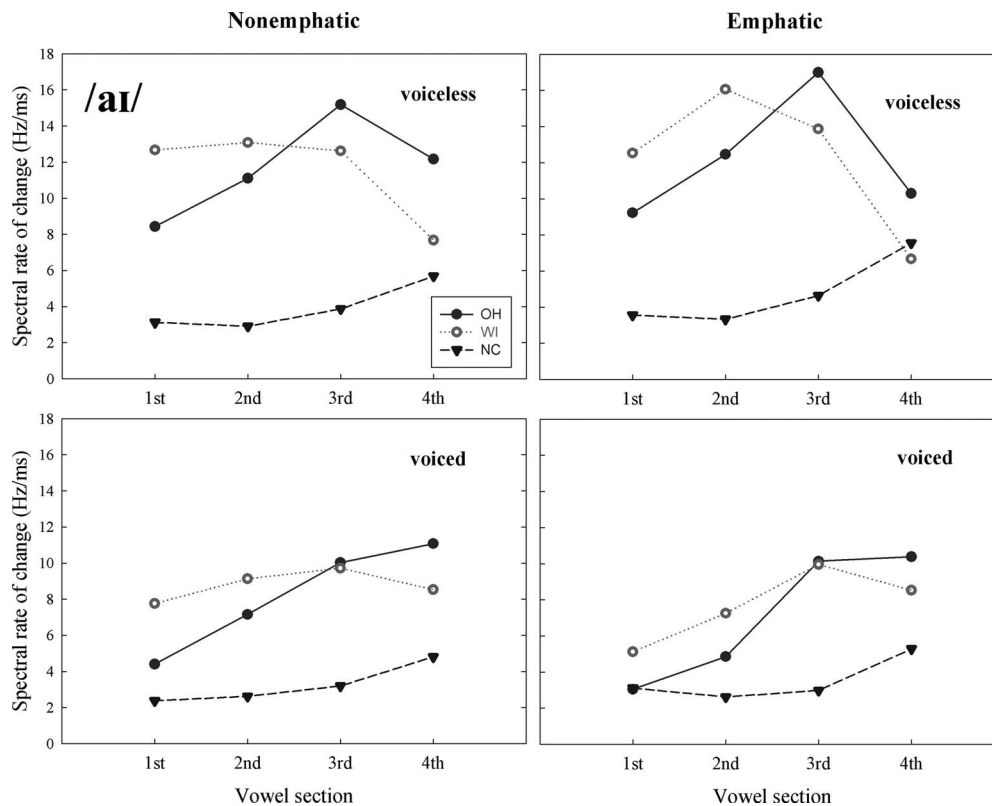


FIG. 8. Mean spectral roc for the diphthong /aɪ/ in the four consecutive sections of the target in nonemphatic (left) and emphatic (right) positions across the dialects.

## B. Dialectal differences in the dynamics of /aɪ/

Regional dialect turns out to be a rich source of phonetic variation. The diphthong /aɪ/ is a good example of how differently vowels can be produced from region to region. This diphthong can be almost monophthongal in North Carolina and it can have two manifestations in even closely related Midwestern dialects spoken in Ohio and Wisconsin. It is interesting to note how the proportional differences in duration and TL arising from the effect of consonantal context interact with dialect-specific variation in spectral roc.

As shown in Table VI, the effects of consonantal context and dialect on the spectral roc of /aɪ/ were particularly strong. From Table VIII we find that while vowel duration was reduced in voiceless context (especially for the Ohio and Wisconsin variants), the TL increased greatly. The shorter duration and greater TL affected the spectral roc which was actually higher in the voiceless than the voiced context (see the negative percentage values). This apparent divergence from the expected pattern of reduction in the voiceless context supports earlier reports in the literature, however. For example, Gay (1968) found that the increased duration of /aɪ/ in *bide* relative to *bite* is accomplished primarily by a lengthening of the steady-state onset of the diphthong while both the gliding portion and the diphthongal offset lengthen to a lesser extent. The lengthening of the onset was verified by Jacewicz *et al.* (2003). This study also showed that the F2 change in *bide* was smaller mostly due to lower terminal frequency values of the offglide. A larger F2 change coupled with a shorter duration means that *bite* will have a higher spectral roc than *bide*. The present results are in accord with

these previous findings and can be easily inferred from the patterns in /aɪ/ displayed in Figs. 1 and 2. That is, there is a smaller spectral change in the first two sections (i.e., over the first three data points) of the Ohio and Wisconsin diphthongs in the voiced context and a comparatively greater formant movement in the voiceless context. Given shorter vowel duration in the latter context, we can thus explain the much higher spectral roc in the voiceless context than in the voiced.

There were also notable dialectal differences in the production of the diphthong. Figure 8 gives a more detailed account of the spectral roc over the four vowel sections. In the voiceless context, the first section of the Wisconsin diphthong shows a higher spectral roc than the Ohio variant and reaches its spectral maximum in the second section while the Ohio /aɪ/ has its maximum later, in the third section. These differences most likely reflect dialect-specific articulatory movements and pace in order to attain the target offglide. In the voiced context, the spectral roc was much lower in general, and the maxima were reached later in the diphthong, in the fourth section of the Ohio vowel and in the third section of the Wisconsin variant. This temporal shift in the spectral roc maximum can be explained on the basis of the lengthening of the diphthongal onglide in the voiced context as discussed above. In general, then, the Wisconsin variant starts with faster articulatory movements which results in the higher roc; the Ohio variant begins with slower movements and reaches higher roc than the Wisconsin variant later in time, closer to its second target /ɪ/.

In sharp contrast to both Midwestern diphthongs is the spectral roc of the North Carolina /aɪ/ which did not change much over the course of its duration and increased only slightly in the fourth section. This result reflects the monophthongal property of /aɪ/ in this variety of American English.

### C. Spectral dynamics and the listener

An obvious question arises as to how sensitive listeners are to these types of spectral changes. Do they recognize dialect-specific spectral variations or are these changes too minor to identify vowels as belonging or not belonging to their own regional variety? The spectral variations examined in this study were limited in terms of the type of consonantal context used and the number of levels of vowel emphasis. Clearly, other contexts will introduce other changes to the dynamic structure of vowels. Will the dialect-specific spectral features persist in these contexts or will they be obscured by contextual variation? These issues are important from yet another perspective, namely, that vowels change their properties across generations as a part of the process known as sound change. Will the dynamic spectral variations in vowels differ between younger and older speakers who grew up in the same dialect area? Will these variations be related to specific vowel shifts and changes currently taking place across geographic regions in the United States such as Northern Cities Shift or Southern Vowel shift? Further research is planned to determine whether and to what extent spectral dynamics contributes to the sound change in progress. The methods presented in this paper may prove useful in these efforts.

### ACKNOWLEDGMENTS

This study was supported by Research Grant No. R01 DC006871 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. The authors would like to thank Joseph Salmons for his contributions to this research. The comments and suggestions of two anonymous reviewers on an earlier version of the paper are greatly appreciated.

<sup>1</sup>In this particular case, the acoustic proximity of both vowels may introduce some perceptual confusion, especially for listeners who grew up in a different dialect area. However, the differences in duration and in the position of the initial portion of each vowel including vowel nucleus (or 50% point) may contribute to a perceptual distinctiveness of both vowels.

<sup>2</sup>We used separate ANOVAs for each vowel rather than a single ANOVA because the main effect of vowel was expected to be significant for the present selection of vowels. Of interest to us were the dialectal differences within each vowel category and not the phonetic differences between vowels.

<sup>3</sup>For the vowels /ɪ, ε, æ/, the context × emphasis interaction arose from the fact that the spectral roc difference between the voiceless emphatic and voiceless nonemphatic contexts was significantly larger than that between voiced emphatic and voiced nonemphatic contexts. For /ɪ/, the mean values were 5.3 and 4.8 Hz/ms for voiceless emphatic vs nonemphatic context as compared to 4.6 and 4.7 Hz/ms for voiced emphatic vs nonemphatic, for /ε/: 5.1 and 4.2 Hz/ms vs 4.2 and 4.3 Hz/ms, respectively, and for /æ/: 4.1 and 3.8 Hz/ms vs 3.8 and 3.8 Hz/ms, respectively. However, the effects of context and emphasis created greater variation in the spectral roc for the diphthongal vowels /e, aɪ/. For /e/, it was the nonemphatic position in which roc was higher, and this was true for both voiceless and voiced contexts (the means were 6.1 and 6.6 Hz/ms for voiceless em-

phatic vs nonemphatic context as compared to 4.3 and 5.3 Hz/ms for voiced emphatic vs nonemphatic). For /aɪ/, there was a mixed pattern of variation in that the spectral roc was higher in the emphatic position for the voiceless context (means=9.8 and 9.1 Hz/ms for emphatic and nonemphatic, respectively) and it was higher in the nonemphatic position for the voiced (means=6.1 and 6.7 Hz/ms for emphatic and nonemphatic, respectively).

<sup>4</sup>Although we used statistical evidence as a guide to the strength of phonetic effects, we acknowledge complications in the interpretation of some of the present results. As shown in Lindblom *et al.* (2007), there is a complex interaction between variation as a function of consonantal context and emphatic stress so that “coarticulatory interactions between the C and V undergo complex, and often subtle, ‘pulls’ and ‘pushes’” (p. 3803). In particular, characterization of locus equation slope (used as an index of degree of coarticulation) may be confounded by the effects of emphatic stress on F2 midpoint values. The effects of emphatic stress on the consonant onset and vowel F2 midpoint need to be separated. In the present study, the effects of emphatic stress may interact with the effects of consonantal context on vowel inherent spectral change in ways difficult to assess given the present design, in that we made only limited systematic modifications of phonetic context. Our current focus was to examine whether the effects of dialect on formant dynamics still persist in the presence of variation coming from the two phonetic sources.

- Adank, P., van Hout, R., and Smits, R. (2004). “An acoustic description of the vowels of Northern and Southern Standard Dutch,” *J. Acoust. Soc. Am.* **116**, 1729–1738.
- Andruski, J. E., and Nearey, T. M. (1992). “On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables,” *J. Acoust. Soc. Am.* **91**, 390–410.
- Byrd, D. (1994). “Relations of sex and dialect to reduction,” *Speech Commun.* **15**, 39–54.
- Clopper, C. G., Pisoni, D., and de Jong, K. (2005). “Acoustic characteristics of the vowel systems of six regional varieties of American English,” *J. Acoust. Soc. Am.* **118**, 1661–1676.
- Ferguson, S. H., and Kewley-Port, D. (2002). “Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **112**, 259–271.
- Fox, R. A. (1983). “Perceptual structure of monophthongs and diphthongs in English,” *Lang Speech* **26**, 21–60.
- Gay, T. (1968). “Effect of speaking rate on diphthong formant movements,” *J. Acoust. Soc. Am.* **44**, 1570–1573.
- Harrington, J. (2006). “An acoustic analysis of ‘happy-tensing’ in the Queen’s Christmas broadcasts,” *J. Phonetics* **34**, 439–457.
- Harrington, J., and Cassidy, S. (1994). “Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English,” *Lang Speech* **37**, 357–373.
- Harrington, J., Kleber, F., and Reubold, U. (2008). “Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study,” *J. Acoust. Soc. Am.* **123**, 2825–2835.
- Hillenbrand, M., and Gayvert, R. T. (1993). “Vowel classification based on fundamental frequency and formant frequencies,” *J. Speech Hear. Res.* **36**, 694–700.
- Hillenbrand, J. M., and Nearey, T. M. (1999). “Identification of resynthesized /hVd/ utterances: Effects of formant contour,” *J. Acoust. Soc. Am.* **105**, 3509–3523.
- Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (2001). “Effects of consonantal environment on vowel formant patterns,” *J. Acoust. Soc. Am.* **109**, 748–763.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). “Acoustic characteristics of American English vowels,” *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Jacewicz, E., Fox, R. A., and Salmons, J. (2007). “Vowel duration in three American English dialects,” *Am. Speech* **82**, 367–385.
- Jacewicz, E., Fox, R. A., O’Neill, K., and Salmons, J. (2009). “Articulation rate across dialect, age, and gender,” *Lang. Var. Change* **21**, 233–256.
- Jacewicz, E., Fujimura, O., and Fox, R. A. (2003). “Dynamics in diphthong perception,” in *Proceedings of the XVth International Congress of Phonetic Sciences*, edited by M. J. Solé, D. Recasens and J. Romero, Barcelona, Spain, pp. 993–996.
- Jenkins, J. J., Strange, W., and Edman, T. (1983). “Identification of vowels in ‘vowelless’ syllables,” *Percept. Psychophys.* **34**, 441–450.
- Jenkins, J. J., Strange, W., and Miranda, S. (1994). “Vowel identification in mixed-speaker silent-center syllables,” *J. Acoust. Soc. Am.* **95**, 1030–

- Kewley-Port, D., and Neel, A. (1996). "Perception of dynamic properties of speech: Peripheral and central processes," in *Listening to Speech: An Auditory Perspective*, edited by S. Greenberg and W. A. Aisnworth (Lawrence Erlbaum Associates, London), pp. 49–61.
- Labov, W., Ash, S., and Boberg, C. (2006). *Atlas of North American English: Phonetics, Phonology, and Sound Change* (Mouton de Gruyter, Berlin).
- Lehiste, I., and Peterson, G. (1961). "Transitions glides, and diphthongs," *J. Acoust. Soc. Am.* **33**, 268–277.
- Lindblom, B. (1963). "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773–1781.
- Lindblom, B., Agwuele, A., Sussman, H. M., and Cortes, E. E. (2007). "The effect of emphatic stress on consonant vowel coarticulation," *J. Acoust. Soc. Am.* **121**, 3802–3813.
- Milenkovic, P. (2003). TF32 software program, University of Wisconsin, Madison, WI.
- Moon, S.-J., and Lindblom, B. (1994). "Interaction between duration, context, and speaking style in English stressed vowels," *J. Acoust. Soc. Am.* **96**, 40–55.
- Morrison, G. S. (2009). "Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs," *J. Acoust. Soc. Am.* **125**, 2387–2397.
- Nearey, T. M., and Assmann, P. F. (1986). "Modeling the role of inherent spectral change in vowel identification," *J. Acoust. Soc. Am.* **80**, 1297–1308.
- Olive, J. P., Greenwood, A., and Coleman, J. (1993). *Acoustics of American English Speech* (Springer-Verlag, New York).
- Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," *J. Acoust. Soc. Am.* **32**, 693–703.
- Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (Allyn and Bacon, Boston, MA).
- Strange, W. (1987). "Information for vowels in formant transitions," *J. Mem. Lang.* **26**, 550–557.
- Strange, W. (1989). "Dynamic specification of coarticulated vowels spoken in sentence context," *J. Acoust. Soc. Am.* **85**, 2135–2153.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.* **74**, 695–705.
- Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (1976). "Consonant environment specifies vowel identity," *J. Acoust. Soc. Am.* **60**, 213–224.
- Van Son, R. J. J. H., and Pols, L. C. W. (1992). "Formant movements of Dutch vowels in a text, read at normal and fast rate," *J. Acoust. Soc. Am.* **92**, 121–127.
- Watson, C. I., and Harrington, J. (1999). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," *J. Acoust. Soc. Am.* **106**, 458–468.
- Wouters, J., and Macon, M. W. (2002a). "Effects of prosodic factors on spectral dynamics, I. Analysis," *J. Acoust. Soc. Am.* **111**, 417–427.
- Wouters, J., and Macon, M. W. (2002b). "Effects of prosodic factors on spectral dynamics, II. Synthesis," *J. Acoust. Soc. Am.* **111**, 428–438.
- Zahorian, S. A., and Jagharghi, A. (1993). "Spectral-shape features versus formants as acoustic correlates for vowels," *J. Acoust. Soc. Am.* **94**, 1966–1982.