Original Paper

Phonetica 2018;75:273–309
DOI: 10.1159/000484610

Received: December 12, 2016
Accepted after revision: October 4, 2017

Regional Variation in Fundamental Frequency of American English Vowels

Ewa Jacewicz Robert Allen Fox

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

Abstract

We examined whether the fundamental frequency (f0) of vowels is influenced by regional variation, aiming to (1) establish how the relationship between vowel height and f0 ("intrinsic f0") is utilized in regional vowel systems and (2) determine whether regional varieties differ in their implementation of the effects of phonetic context on f0 variations. An extended set of acoustic measures explored f0 in vowels in isolated tokens (experiment 1) and in connected speech (experiment 2) from 36 women representing 3 different varieties of American English. Regional differences were found in f0 shape in isolated tokens, in the magnitude of intrinsic f0 difference between high and low vowels, in the nature of f0 contours in stressed vowels, and in the completion of f0 contours in the context of coda voicing. Regional varieties utilize f0 control in vowels in different ways, including regional f0 ranges and variation in f0 shape.

Introduction

Acoustic characteristics of American English vowels have been studied systematically since Peterson and Barney (1952), with formant patterns and spectral dynamics being the primary focus of investigation (e.g., Fox and Jacewicz, 2009; Hillenbrand et al., 1995, 2001). While amplitude variation has recently been explored in greater detail (Jacewicz and Fox, 2008a) and the importance of vowel duration has long been recognized (Hillenbrand et al., 2000; Lehiste and Peterson, 1960), fundamental frequency (f0) has received relatively limited attention in much of the work on vowel acoustics, sometimes being included for completeness of description rather than receiving more fundamental consideration such as found in early work by House and Fairbanks (1953) or Lehiste and Peterson (1961). The "secondary" status of f0 as a supporting acoustic cue to vowel identity has been well maintained since then, owing by and large to a lack of serious interest in exploring f0 variations on the one hand and to methodological issues related to how to best measure f0 on the other.

KARGER

E-Mail karger@karger.com
www.karger.com/pho

Ewa Jacewicz
Department of Speech and Hearing Science
Ohio State University
1070 Carmack Road, Columbus, OH 43210 (USA)
E-Mail jacewicz.1@osu.edu

After a period of relative neglect, the need for a better understanding of f0 characteristics has resurfaced in the new context of sociophonetics, including regional variation. Work in this area revealed that both the formant patterns and duration of American English vowels vary systematically across geographic regions in the USA (Clopper et al., 2005; Jacewicz et al., 2007, 2011a). Comparatively little, if anything, is known about possible effects of regional variation on f0 in those vowels. The scant existing evidence admits the possibility that f0 variation can be regulated to some extent by regional differences in prosodic patterns (for an overview, see Thomas, 2011). For the latter, differences were reported in the rising pitch accents between Minnesotan and Southern Californian speakers (Arvaniti and Garding, 2007) and in the frequency of pitch accent types and phrasal-boundary tone combinations between Midwestern and Southern speakers (Clopper and Smiljanic, 2011). Although it is still unclear how variations of this type can systematically influence f0 of a vocalic segment, it is certainly possible that different varieties utilize pitch patterns to express specific cultural renditions of such elements as sentential focus or emphasis, of which syllable nuclei become the primary carriers. It is also possible that communities differ in the prevailing pitch ranges that they use for speech. Consequently, speakers of a particular variety may acquire an internal representation of a pitch range in their speech community and this representation, shaped by long-term exposure to the speech of others, may influence their speech, including vowel production. This mental representation may then be utilized when acquiring a particular regional variety (Deutsch et al., 2004, 2009).

More recently, evidence for the sociocultural sources of f0 use was presented in an acoustic investigation of intrinsic f0 in vowels (Jacewicz and Fox, 2015a). That study examined f0 patterns in the high /ɪ/ and the low /æ/, and found that the size of f0 difference between the 2 vowels varied with the regional variety of American English when the vowels were stressed. Any f0 differences diminished, however, when the vowels were unstressed. The results for stressed vowels suggest that speakers can implement different f0 patterns reflecting regional characteristics of speech. It appears that f0 control can be, at least in part, learned separately as a part of a cultural "tune" (Titze, 2000, p. 211) and thus can interact with the intrinsic attributes of different vowel qualities to convey sociocultural identity.

In the current paper, we further inquire into the nature of this cultural influence, testing the general hypothesis that f0 in a vowel varies systematically as a function of regional variation. We emphasize that the current study is designed to address those f0 properties that are attributable to vowel identity, i.e., contribute to the acoustic specification of a vowel (Kent and Read, 1992), and not to broader prosodic structures related to prosodic organization. We will operationally refer to these segmental f0 variations as "vowel f0," a term used in the literature interchangeably with "f0 of a vowel" and "f0 in a vowel" (e.g., Hanson, 2009; Kent and Read, 1992; Steele, 1986). These segmental f0 variations will be controlled for in 2 ways, by investigating f0 of a vocalic segment in syllables produced as isolated items (experiment 1), and in stressed and unstressed syllables in sentential contexts (experiment 2). In using the term "vowel f0" we are primarily concerned with systematic f0 variation of a vocalic segment as a function of the 2 production types, recognizing that any production of natural speech, including citation form syllables, necessarily shares an acoustic form with the use of f0 variation for intonational purposes, and that these 2 interact.

Considering regional variation, it is the case that major vowel changes have split American English into regional varieties. Chain shifts such as the Southern Shift, the

Northern Cities Shift, the Canadian Shift, the Pittsburgh Shift, or the Back Chain Shift, rotating vowel systems are responsible for the increasing divergence of North American English dialect regions (Labov, 2010; Labov et al., 2006). The notion of chain shifting itself originates in descriptive historical phonology and represents sound change, involving "encroachment of one phoneme into the phonological space of another" (Gordon, 2002). In chain shifts, vowels change their positions in interlocking groups. For example, one vowel can "push" another vowel moving it to a different position while still maintaining phonological contrast between the 2 (the so-called push chain).

With the advancement of speech technology, the chain-like vowel rotations in the American English vowel system have been intensely studied by sociolinguists (cf. Labov, 1994), and defined on the basis of 2 primary directional dimensions, the high versus low and the front versus back. The high versus low dimension reflects changes in vowel height measured as changes in the frequency of the first formant (F1). In the sociophonetic context pertaining to chain shifts, F1 change is interpreted as raising (or lowering) of a vowel relative to other vowels in the system. Likewise, the front versus back dimension is commensurate with changes in the frequency of the second formant (F2) that correspond to a relative fronting (advancement) or backing (retraction) of a vowel in the vowel space (e.g., Boberg, 2005; Labov et al., 1972). Admittedly, "the first two formants provide only a first approximation to the characterization of vowel timbre, but one that has proven useful in studies of dozens of sound changes" (Labov, 1994, p. 55).

Of interest to our current investigation is the relationship between vowel height (F1) and vowel f0, on the premise that if regional vowel rotations involve positional changes in F1, they will also influence f0 of those vowels. This reasoning is based on the assumption that f0 covaries with F1 so that high vowels (with low F1) have high f0, and low vowels (with high F1) have low f0, a finding dating back at least to House and Fairbanks (1953). It needs to be pointed out that the exact causes of this "intrinsic f0" are still debated and different hypotheses diverge primarily in 3 respects, positing that intrinsic f0 (1) may reflect an automatic (passive) aspect of vowel articulation without being intended itself (Honda, 1983; Honda et al., 1999; Ohala and Eukel, 1987; Sapir, 1989; Whalen and Levitt, 1995), (2) may be actively and deliberately controlled by the speaker to perceptually enhance vowel category distinctions along the high versus low dimension (Diehl and Kluender, 1989; Kingston, 1992, 2007), (3) may be physiological in nature but it may also be, to some extent, enhanced by the speaker to improve certain aspects of vowel characteristics (Honda and Fujimura, 1991; Van Hoof and Verhoeven, 2011).

For present purposes, it is important to consider what predictions these 3 models could potentially make regarding the effects of chain-like changes involving lowering or raising of neighboring vowels in the vowel space on corresponding changes in f0. We begin by considering predictions from the first model. Accordingly, if the F1/f0 relationship is automatic, then a change in F1 will necessarily involve a change in f0. However, the F1/f0 relationship was studied primarily in high and low vowels (Ladd and Silverman, 1984; Shadle, 1985; Steele, 1986; Whalen and Levitt, 1995), and it remains to be shown whether the relationship actually holds for neighboring vowels of phonologically different height that are less spectrally distant, such as /ɛ/ and /ɪ/ that are often involved in regional chain shifts. To that end, Fischer-Jørgensen's (1990) findings for German, including mismatches between F1 and f0 in /ɪ/ and /e/, certainly suggest that biomechanical coupling is not the only basis for the "intrinsic" aspect of f0 and compensatory mechanisms could also be involved. In terms of the second and third models, some controlled muscular actions on the part of the speakers, especially in the context of regional variation, cannot be discounted. For example, social attitudes and cultural identity have always played a role in the maintenance of (and divergence from) regional varieties, which underscores an active involvement of the speakers in the production of locally meaningful pronunciation variants and their transmission across generations (e.g., Kurath, 1939; Stanford et al., 2012).

Given the unsettled status quo as to the source and nature of f0 covariance with F1, we emphasize that predictions about f0 based on regional variation and change in F1 can be cumbersome not only because of the complexity of interaction between the control of f0, the configuration of the vocal tract, and the speaker, but also due to methodological choices. As we outline below, a careful selection of the most effective measures becomes a pivotal issue in addressing questions related to the sociophonetic relevance of the F1/f0 relationship.

The Relationship between Vowel Height and f0

Evidence for the existence of intrinsic f0 in vowels comes primarily from experiments using isolated syllables, isolated words and, less frequently, syllables or words embedded in carrier sentences (Ladd and Silverman, 1984; Shadle, 1985; Steele, 1986; Whalen and Levitt, 1995). A great number out of the 58 published sources compiled in Whalen and Levitt (1995) did not report where in the time course of a vowel f0 was measured. The most common approaches that can be found in the literature include averaging f0 over the course of a vowel (House and Fairbanks, 1953), averaging f0 over a defined "steady state" (Hillenbrand et al., 1995), sampling at peak f0 when the syllable is stressed (Lehiste and Peterson, 1961), and sampling f0 at the vowel's midpoint (e.g., Leung et al., 2016). In another study, Whalen et al. (1998) measured f0 in /ɑ, i, u/ vowels produced in isolation, "near the beginning of the vowel, before the f0 fall began" (p. 130). Shadle (1985) still used a different approach, averaging frequencies of 3 pitch periods centered at a temporal location 50 ms prior to the [b] closure of the syllable coda (the main experiment, p. 1564).

The measurement location becomes far from trivial in the context of regional variation. For example, the raising and lowering of vowels in the vowel space have been typically determined on the basis of the vowel's nucleus defined as a steady state in both formants (Labov et al., 2006). However, as pointed out in Labov et al. (2006, p. 38), vowels in some regional varieties may demonstrate "a rise and fall in F1, with a maximal value of F1 representing the lowest point reached by the tongue," and thus measurements obtained at this location are important in characterizing directionality and magnitude of positional change in the vowel space. Moreover, in regional variation, diphthongal glides may be as important as vowel nuclei since monophthongization and diphthongization of a vowel can also mark a distinction between varieties or sound change, of which the monophthongization of /aɪ/ in Southern American English is the most famous example. An additional complication comes from the finding that even the nominal monophthongs can display a significant amount of F1 change over the course of a vowel and regional varieties also differ in the nature of this dynamic specification (Fox and Jacewicz, 2009; Jacewicz et al., 2011a). While quantitative analyses using curve-fitting parameterizations (e.g., Morrison and Assmann, 2013; Risdal and Kohn, 2014; Watson and Harrington, 1999) and smoothing spline analysis of variance (ANOVA) (Koops, 2010) have been generally successful in modeling formant change, the corresponding functional fits to f0 trajectories were not included in those models. How, then, can the F1/f0 relationship be established in the context of regional variation?

The major methodological challenge is in associating a particular F1 value as an indicator of vowel position in the space with the corresponding f0. To that end, we can expect considerable mismatches such as that the location of the vowel nucleus (averaged over a steady state in F1) may correspond to the f0 fall in a vowel rather than reflect a more optimal f0 value obtained closer to the vowel's onset (Whalen et al., 1998). On the other hand, if f0 is to be averaged over the entire duration of a vowel, this average f0 value may not properly reflect the positional variation in vowel height determined on the basis of the vowel's nucleus (Labov et al., 2006) because regional varieties may differ in the magnitude of the f0 fall. Consequently, a greater (or smaller) f0 fall may "artificially" lower (or raise) the f0 value in relation to F1, leading to misinterpretations of the F1/f0 relationship. Still to the point, how can we determine f0 change in a diphthong? Should 2 f0 measurements be obtained at 2 F1 locations assuming that f0 values will correspond to the 2 intended targets rather than exhibit an expected f0 fall over the diphthong's duration? The existing literature has not addressed this question thus far. With these issues in mind, experiment 1 was aimed at illuminating possible effects of regional variation on intrinsic f0 of vowels that are either involved or not involved in regional chain shifts. To determine the "baseline" for each regional variety, the vowels in experiment 1 were produced as isolated items in hVd context.

The Prosodic Effects on f0 in Vocalic Segments

A recognized complicating factor in estimating f0 in vowels is that the F1/f0 relationship may be obscured by the linguistic and indexical properties that influence phonation, emotion, speaking style and, predominantly, prosody. In fact, it has been argued that intrinsic f0 is an artificial effect of high experimental control in carrier sentences and that it is mitigated in fluent reading (Umeda, 1981). Testing this possibility, Ladd and Silverman (1984) found, however, that intrinsic f0 was still present in connected speech but the magnitude of the f0 difference between high and low vowels was significantly smaller when compared with that in carrier phrases. Of current interest is whether any effects of regional variation on f0 patterns identified in tokens produced in isolation (experiment 1) will also persist in vowels in connected speech (experiment 2).

As can be expected, intrinsic f0 in connected speech is further regulated by stress pattern and sentence prosody. As shown previously, the effects of regional variation on intrinsic f0 variations in stressed syllables can be strong and significant, and are mostly diminished in unstressed syllables (Jacewicz and Fox, 2015a). However, the timing of pitch accent, the melodic feature that falls on the most prominent stressed syllable in a sentence, can also reflect differences in the regional use of prosody. As demonstrated in a large body of work, particularly within the framework of intonational phonology (Beckman and Pierrehumbert, 1986; Ladd, 1996), the time course of the f0 contour interacts systematically with prosodic context such as phrase-level intonation. More recently, work in this area produced evidence for the distinctive use of f0 patterns in regional varieties of the same language, including pitch peak alignment in Serbian and Croatian (Smiljanic, 2006) and in 2 varieties of British English (Ladd et al., 2009). We can thus expect that at least some regional varieties of American English may differ in how peak placement is related to prosodic and temporal structure, such as occurring earlier in time when a syllable is lengthened by an upcoming word boundary, an effect induced by prosodic context (Silverman and Pierrehumbert, 1990). The variation in peak placement may be related to regional differences in mapping between phonological categories of representation and f0 peak alignments (Pierrehumbert and Steele, 1989). For example, shifting an f0 peak through a weak-strong or strong-weak syllable may be related to 2 categories of representation rather than 1 (Redi, 2003), and regional varieties may differ in how local maxima, rises, and falls contained in intonation contours are related to the segmental string.

However, interactions between prosodic phonological representation, phonetic variation in peak or valley locations, and regional variation in intrinsic f0 can be very complex, which necessitates a separate investigation, far beyond the primary phonetic focus of the current study. In the context of this paper, we aim to determine whether regional varieties can differ in how they implement the effects of *phonetic segmental context* on f0 variation in vowels in connected speech. The specific contextual effect examined in experiment 2 is that of obstruent consonant voicing in syllable coda on the f0 of the preceding vowel. The relevant background literature that led us to select this particular variable is discussed in greater detail below.

The Effects of Obstruent Consonant Voicing on Vowel f0

Contextual effects on f0 in American English vowels were first reported by House and Fairbanks (1953) who found that the average f0 of vowels produced between 2 symmetrical voiceless consonants (C_1VC_1) was higher than in a context of symmetrical voiced consonants. A closer investigation of these segmental effects in this and other studies revealed that the effect of the voiced/voiceless dichotomy occurred primarily in the early part of a vowel, near its onset (Hombert et al., 1979; Ohde, 1984; see also Hanson, 2009, for a review). These f0 variations can be explained as arising from a carryover effect of glottal stiffening or slackening gestures involved in the production of the preceding consonants that cause the vocal folds to vibrate faster in vowels following a voiceless obstruent than in vowels following a voiced obstruent (Halle and Stevens, 1971; Löfqvist et al., 1989). This effect can persist well into the following vowel. For example, Lehiste and Peterson (1961) also found it at peak f0 in stressed vowels (at “the peak of the intonational contour”) that was again higher when the vowel was preceded by a voiceless consonant than when the consonant was voiced. A recent detailed study of this pattern in languages other than English also found the same co-occurrence in French and Italian, suggesting that the pattern is fundamentally phonetic, arising from common articulatory adjustments to inhibit phonation (Kirby and Ladd, 2016).

Several early studies also reported a similar voiced/voiceless dichotomy with regard to postvocalic consonants, with voiceless consonants raising the f0 of the preceding vowels and voiced consonants lowering their f0 (Mohr, 1968; Slis, 1966), although more variable patterns have also been found (Lea, 1973). The lack of consistency for the effects of consonant voicing on the preceding vowel may be, in part, related to variable timing patterns within the syllable that affect vowel duration and thus terminal f0 values. In particular, syllable-final voiceless obstruents in English shorten the duration of the preceding vowel. If the syllable is accented or produced with a greater prominence in general, it may happen that the tonal contour of the vowel is “clipped” (or cut short) and the terminal f0 value is higher before a voiceless obstruent than before a voiced obstruent (see Hombert et al., 1979, for further discussion). The effects of the postvocalic obstruent voicing have not yet been systematically investigated, however, which is a notable gap in the literature pertaining to variation in vowel f0.

Addressing this gap, obstruent voicing in experiment 2 was varied in the postvocalic position. This choice was made, in part, due to availability of the production data that were already collected for a larger related project. However, irrespective of this constraint, we reasoned that voicing manipulations of the obstruents that follow (rather than precede) a vowel are likely to manifest differential dialect effects on f0 patterns. This is because not only durations of vowels can differ as a function of regional variation (Jacewicz et al., 2007) but different varieties may also differ in their realization of the tonal contour in vowels produced with greater emphasis such as in words carrying semantic focus. If so, then the prolonged duration of an emphatic vowel before a voiced plosive will likely afford an additional time for a completion of the f0 contour in a manner specific to a particular variety. However, a shortened duration before a voiceless stop may (1) “clip” the f0 contour and thus reduce the regional differences because the raised terminal f0 values are now comparable or (2) maintain the regional differences by preserving the f0 contour shapes that appeared in the voiced context. A third possibility also exists: some varieties may “clip” and some may maintain the same f0 contour shape in both contexts. These possibilities will be tested in the current analysis of vowel f0 in 3 regional varieties of American English.

The Study: Selection of Regional Varieties and Participants

Speech samples for the 2 experiments were selected from a large cross-dialectal corpus collected to elicit variable vowel productions across several generations of speakers. In the current work, we utilized a subset of this speech material, which was previously used to explore dynamic formant patterns in vowels (Fox and Jacewicz, 2009; Jacewicz et al., 2011a). We refer interested readers to these studies for relevant descriptions pertaining to spectral characteristics of the vowels examined here.

Selection of Regional Varieties

Our previous work found significant regional differences in the spectral and temporal structures of vowels (Fox and Jacewicz, 2009; Jacewicz et al., 2007, 2011a) and consonants (Jacewicz et al., 2009). In all these studies, Southern American English presented a distinct acoustic profile when compared with more general American English varieties spoken in the North. The distinctive acoustic characteristics of Southern American English vowels including their greater formant movement and longer durations were perceptually salient (Jacewicz and Fox, 2012) and, very possibly, contributed to an intelligibility benefit of the Southern speech in noisy conditions (Jacewicz and Fox, 2015b). In the current study, we turn to the f0 in vowels in the same Southern American English variety and compare the acoustic results with those for the same 2 Northern varieties studied earlier. We expect f0 variations in the Southern vowels to be somewhat distinct due to widely observed exaggerated pitch rises in stressed syllables that, together with specific type of spectral dynamics, produce the percept of the stereotypical “Southern drawl” (Sledd, 1966). There is no experimental evidence, however, to bear on the question of whether the distinct melodic feature also appears in other than stressed contexts.

The particular Southern variety studied here is the variety spoken in the western part of North Carolina (Jackson County), identified as Inland South on the geographic map of regional variation in American English (Labov et al., 2006). Unlike the remaining regions in the state of North Carolina including Piedmont, Coastal Plain, and Outer Banks, the Appalachian variety in the Inland South is associated with an advanced stage of the Southern Shift in the vowel system, at least in the speech of older residents of the area. The particular spectral characteristics related to the Southern Shift include monophthongization of /aɪ/ and positional reversals of front vowels /e/ and /ɛ/, and /i/ and /ɪ/, mostly due to the raising and fronting of the 2 lax vowels /ɛ/ and /ɪ/. As a whole, the vowel system of the Inland South also exhibits a distinct pattern of formant dynamics (Jacewicz et al., 2011a). For practical reasons and to facilitate descriptions of graphical displays, we refer to this variety in the current paper simply as “NC” bearing in mind that the reference does not include other regional varieties spoken in the state of North Carolina, where the Southern Shift has steadily been in recession for several decades.

The second regional variety under study is that spoken in Southeastern Wisconsin, which is regarded as part of the Inland North region, at the western edge of the Northern Cities Shift area (Labov et al., 2006). The vowel system of Southeastern Wisconsin demonstrates several features of the Northern Cities Shift, including a relatively raised variant of /æ/ and lowering and centralization of /ɛ/. Notably, the positions of both /u/ and /o/ are far back in the vowel space, and both vowels are monophthongal, which is in sharp contrast with the fronted /u/ and the diphthongized /o/ in the Western NC variety. Overall, the 2 regional vowel systems differ greatly in relative positions of vowels in the vowel space and in the nature and amount of formant dynamics. We will refer to the Southeastern Wisconsin variety as “WI” throughout the paper.

Finally, the third regional variety spoken in the central part of Ohio was selected as representing yet different patterns of both positional variation and spectral changes. Although not considered to be affected by any known chain shift for long, recent studies indicate that parts of the vowel system in Central Ohio are undergoing a set of changes, including lowering of /æ/ and raising of /ɑ/ that results in a low back merger of /ɑ/ and /ɔ/ in the speech of younger generations (Durian et al., 2010; Jacewicz et al., 2011a). Overall, this variety displays a combination of spectral characteristics of vowels of the North and of the South, including the fronted /u/. Technically, the Central Ohio variety is regarded as part of the Midlands dialect region (Labov et al., 2006). We will refer to the Central Ohio variety as “OH” throughout the paper. Detailed graphical displays of vowel spaces in relation to formant dynamics in each of the 3 regional varieties can be found in Fox and Jacewicz (2017).

Participants

Productions of 36 women aged 51–65 years were utilized in the current study. The participants were born, raised, and resided in 1 of 3 geographic dialect areas in the USA: 12 were from Western North Carolina (Jackson County), 12 were from Central Ohio (Columbus and suburbs), and 12 were from Southeastern Wisconsin (Madison area). These participants were a subset of 48 speakers used in Fox and Jacewicz (2009) for vowel formant analysis. The selection criteria for the current f0 study excluded individuals with colds and other respiratory infections, and those with extensive creaky voice productions. As is well known, the above conditions can be detrimental to the accuracy of f0 measurements. Based on spectral and temporal analysis of vowels from several generations of speakers in each of the 3 dialect regions (Jacewicz et al., 2011a), the middle-aged participants were selected to represent the most typical vowel characteristics in each regional variety. The mean age in years (and standard deviation) for NC participants was 56.5 (3.2), for OH it was 57.6 (5.9), and for WI 58.2 (4.0). These slight age differences among the groups were not significant using a 1-way analysis of variance.

Experiment 1: Intrinsic f0 in Vowels Produced in Isolated hVd Tokens

Experiment 1 had 2 goals. The first goal was to determine, for each regional variety, the relationship between vowel height (F1) and f0 as a function of temporal location in a vowel. The second goal was to verify the previous finding that the magnitude of the intrinsic f0 difference between high and low vowels is influenced by regional variation (Jacewicz and Fox, 2015a) using a different set of vowels and different speech materials.

Speech Materials

The stimulus set consisted of 14 vowels /i, ɪ, e, ɛ, æ, ɑ, ɔ, o, u, ʊ, oɪ, aɪ, aʊ, ɝ/ in hVd context, yielding the tokens *heed, hid, heyd, head, had, hod, hawed, hoed, who'd, hood, hoyd, hide, howed*, and *heard*, respectively.

Procedure

The study was conducted simultaneously at each of the 3 testing locations in North Carolina, Ohio, and Wisconsin by the same designated female experimenter at each testing site. Each participant was tested individually, seated in front of a computer monitor. A head-mounted Shure SM10A dynamic microphone was positioned about 1.5 inches from the speaker's lips. The hVd prompts appeared in random order on the screen and the speaker read each item, one at a time. Three repetitions of each item were obtained for a total of 42 tokens from each participant (n = 1,512). No specific instructions were given as to the way the tokens were to be produced. The utterances were recorded and digitized at a 44.1-kHz sampling rate with 16-bit quantization directly onto a hard disk drive. The experiment was under computer control using a custom program written in MATLAB (MathWorks Inc.). The recording levels were controlled for each utterance for each speaker.

Acoustic Analysis

Prior to acoustic analysis, the tokens were digitally filtered and downsampled to 11.025 kHz. The F1 and f0 frequencies were sampled at 5 temporal locations over the course of a vowel corresponding to the 20–35–50–65–80% points. The measurements of vowel onsets and offsets (which were located by hand) served as input to calculations of vowel duration. F1 values were extracted automatically using a custom program written in MATLAB (MathWorks, 2014a). Using the MATLAB built-in functions available for signal processing and graphical display, the program displayed the formant frequency values along with both fast Fourier transform (FFT) and linear predictive coding (LPC) spectra, and a wide-band spectrogram. An autocorrelation linear prediction algorithm with 14 coefficients was used to produce an estimate of the spectral envelope. A 25-ms Hanning window was centered at each temporal point. The corresponding f0 values were obtained using MATLAB's autocorrelation function.

Two reliability checks, one for each measure, were performed on all measurements by 2 different researchers. All automatically extracted formants were checked using the commercially available software package TF32 (Milenkovic, 2003). In addition to the overlay of the LPC on the FFT spectra and LPC formant tracking for visual inspection, the TF32 allows the manual adjustment of analysis parameters, including the analysis bandwidth of the FFT and LPC filter order. While the assumed rule of thumb for choosing the order of the LPC in MATLAB generally protected against missing a formant or merging 2 formants in some of the back vowels, spurious peaks in the estimated spectrum did occur. These spurious peaks were removed manually on the basis of a global visual inspection approach, using simultaneously the displays in the MATLAB window and, on a second screen, in TF32. Manipulations to analysis parameters most often included decreasing the number of LPC coefficients and increasing the analysis bandwidth of the FFT spectrogram in speakers with higher speaking f0. Adjustments of this type using TF32 have been shown to improve formant-frequency measurements for female adults, outperforming several other speech analysis software packages (Derdemezis et al., 2016). Any disagreements in the analysis were resolved between the 2 researchers, and the final formant measurement check (a custom routine written in MATLAB) was performed on all measurements to ensure that the hand-corrected values, displayed as vertical lines in the LPC spectrum overlaid on the FFT spectrum (formant frequency values were also graphically displayed and matched to a wide-band spectrogram), reflected the intended corrections.

All f0 measurements were checked by running a cepstral analysis on all tokens (also implemented in MATLAB) to detect any major discrepancies. Hand correction, if necessary, was done on the basis of f0 tracks generated in TF32 and PitchWorks (Scicon R and D, 2005) analysis software. Altogether, 7,560 F1/f0 measurements were obtained for use in statistical analyses (1,512 vowel tokens × 5 time points).

Results

The Relationship between Vowel Height (F1) and f0

The F1/f0 relationship at each of the 5 time points in a vowel is shown in Figure 1. The plots include all 14 vowels and diphthongs, displayed separately for each regional variety. Regional variation in F1 of individual vowel categories is apparent, reflecting differences in their relative positions within each regional vowel system. It is also of note that the current f0 values tend to be slightly lower when compared with measurements for American English such as in Hillenbrand et al. (1995). These lower f0 values are most likely due to a decrease in speaking f0 as a function of aging (Reubold et al., 2010), a change particularly affecting women (e.g., Nishio and Niimi, 2008; Xue et al., 2008), and our female participants fall into this age range.

A simple linear regression model was used to determine whether f0 could be predicted from F1. The results show that, for each regional variety and for each temporal location, a significant proportion of the variance in f0 was predicted by vowel height. The relationship between F1 and f0 was strong and significant so that as F1 values increased (i.e., vowel height decreased), f0 values decreased. However, there were also notable differences among the regional varieties, to which we now turn.

For the NC variety, F1 was the best predictor of f0 when measured close to the vowel's onset (20% point). However, the relationship between the 2 variables remained

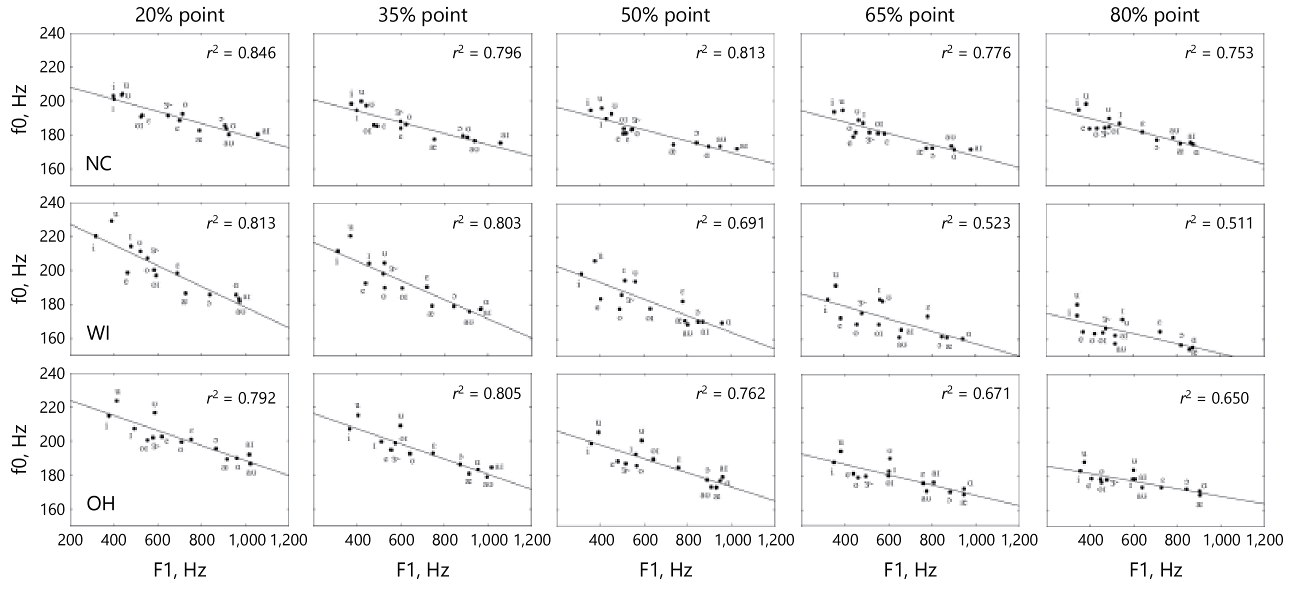


Fig. 1. Plots of f0 by F1 for each vowel category (14 monophthongs and diphthongs) produced in isolated hVd tokens plotted separately for each regional variety: North Carolina (NC), Wisconsin (WI), Ohio (OH). The F1 and f0 measurements were obtained at 5 equidistant time points in each vowel (20–35–50–65–80%). Each data point represents an average of 36 productions (3 repetitions × 12 speakers).

Table 1. Slopes for the regression line of f0 and F1 for measurements at 5 temporal locations in the vowel

Regional variety	20%		35%		50%		65%		80%	
	slope	*r*	slope	*r*	slope	*r*	slope	*r*	slope	*r*
NC	–0.04	0.92***	–0.03	0.89***	–0.03	0.90***	–0.03	0.88***	–0.03	0.87***
WI	–0.06	0.90***	–0.06	0.90***	–0.05	0.83***	–0.04	0.72**	–0.03	0.72**
OH	–0.04	0.89***	–0.05	0.90***	–0.04	0.87***	–0.03	0.82***	–0.02	0.81***

Data are for 14 vowels and diphthongs, averaged over 12 female speakers in each regional variety: North Carolina (NC), Wisconsin (WI), Ohio (OH). ** $p < 0.01$, *** $p < 0.001$.

strong for all temporal locations; as shown in Table 1, all *r* values were greater than 0.8 and significant at $p < 0.001$. The proportion of the explained variance in f0 remained high throughout, decreasing only from 0.846 at the onset to 0.753 at the offset. Importantly, the steepness of the slope was relatively steady for all temporal points, so that with every 1-Hz increase in F1, f0 decreased by approximately 0.033 Hz. These results indicate that, once f0 was "set" at the vowel's onset, it remained relatively unchanged in the course of the vowel's duration.

A different pattern emerged for the WI variant. Consistent with NC, F1 was the best predictor of f0 at the beginning of the vowel (at the 20% point), accounting for 0.813 of the variance in f0. The strong relationship between the 2 variables began to weaken at the midpoint, although 0.691 of the variance in f0 could still be explained by F1. The F1/f0 relationship was further reduced in magnitude when measured later (at the 65% point), where only 0.523 of variance in f0 could be explained by F1 and the corresponding slope became shallower (the reduction in the magnitude of the f0 rate of change was from 0.049, $p < 0.001$ to 0.037, $p = 0.003$). These results indicate a significant change in the F1/f0 relationship in a later part of a vowel, after the midpoint, which was associated with a substantial f0 fall. This overall pattern is thus different from that in NC. It is also important to note that, compared with NC, the WI regression line slope was almost twice as steep when measured at the vowel's onset and approached the shallower NC slope when measured closer to the offset.

The results for OH present yet another pattern. Consistent with WI, F1 was the best predictor of f0 when measured before the midpoint, accounting for approximately 0.80 of the variance in f0. F1 was also a comparatively greater predictor at the offset, being able to explain 0.65 of the variance in f0. Unlike for WI, however, the F1/f0 relationship remained strong across all measurement locations, which was consistent with NC (all *r* values were greater than 0.8 and significant at $p < 0.001$). In terms of the slope, the OH variant was "in between" NC and WI (the OH slope was steeper than that of NC but shallower than that of WI) when measured up to the midpoint, and was the shallowest from all 3 regional varieties at the offset.

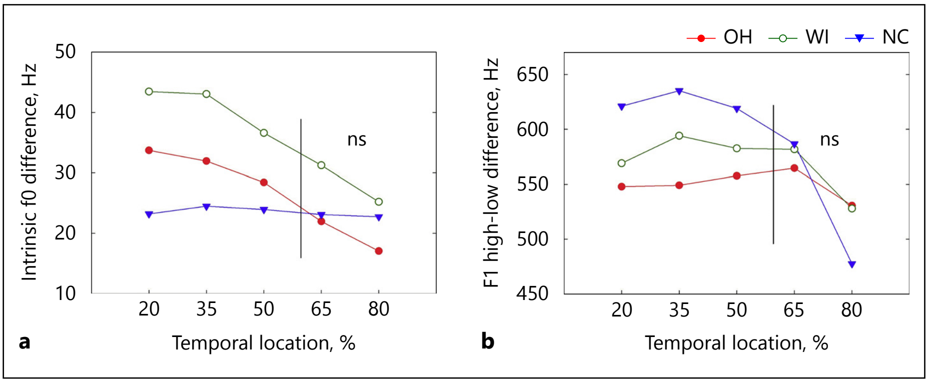
In summary, the results so far strongly support the F1/f0 relationship. However, systematic regional differences were found in the steepness of the slope, which indicates that the rate of f0 change throughout the vowel was variable among regional varieties. This variation most likely reflects both the magnitude of the f0 range utilized by each regional variety and the differential control of the f0 contour. Of note here is the unchanged position of the monophthongal variant of /aɪ/ in the 80% location for NC speakers, whereas both OH and WI speakers show a diphthongal offglide with an increased f0, manifested as a leftward shift along the regression line. We now turn to the second goal of experiment 1, which was to verify that the utilized f0 range (measured as the intrinsic f0 difference between high and low vowels) was significantly influenced by regional variation.

The Intrinsic f0 Difference between High and Low Vowels

Our selection of a high and a low vowel was guided by the following criteria. We selected the vowel /u/ in *who'd* as representing the high vowel because, for all 3 regional varieties, /u/ had a consistently higher f0 than /i/, a trend also found in Whalen and Levitt (1995). Consequently, /u/ could be defined as the vowel with the highest possible f0. In terms of F1, the differences between /i/ and /u/ were relatively small; averaged across all 3 varieties, the F1 of /u/ was about 40 Hz higher than that of /i/, which is consistent with the differences reported in de Boer (2011). The selection of a low vowel was more challenging because 3 vowels, /a, ai, au/ (phonetically: [ɑ, ɑɪ, ɑʊ]), tended to overlap in F1, and diphthongal changes were not manifested until later in a vowel, typically starting at the 65% point. We decided to select the lowest *monophthongal* vowel for each regional variety to ensure that this same vowel would remain "low" across all 5 temporal points. For WI and OH, the [ɑ] in *hod* was the lowest monophthong. For NC, the [ɑɪ] was selected instead because the vowel is produced as a monophthong in this regional variety (*hide* is pronounced as *ha:d*) and, based on F1 values, the [ɑɪ] (and not [ɑ]) represented the lowest monophthong.

The f0 differences between these high and low vowels (henceforth the intrinsic f0 difference) were analyzed using a repeated-measures ANOVA with the within-subject factor temporal location (5 levels) and the between-subject factor regional variety. The ANOVA assumptions were met in this balanced and fully crossed design. The residuals were not considerably skewed, and homogeneity of variance was not violated as determined by Levene's test, $p > 0.05$ (with values ranging from 0.278 to 0.814); also, there were no missing data. As can be expected, the main effect of temporal location was significant [$F(4, 132) = 13.1, p < 0.001, \eta^2_p = 0.285$] indicating that the magnitude of the intrinsic f0 difference was variable across temporal points, descending steadily from onset to offset. Importantly, the main effect of regional variety was significant [$F(1, 33) = 3.6, p = 0.039, \eta^2_p = 0.179$]. Subsequent post hoc tests (Tukey HSD) indicated that the intrinsic f0 difference for NC (mean = 23.5 Hz) was significantly smaller than that for WI (mean = 35.9 Hz), with $p = 0.038$, but OH (mean = 26.6 Hz) did not differ significantly from any other variety. There was also a significant interaction between these 2 factors [$F(8, 132) = 2.8, p = 0.007, \eta^2_p = 0.144$], which is visualized in Figure 2a. Exploring this interaction, an independent *t* test indicated that the differences between NC and WI were significant only for the first 3 temporal points and were not significant for either the 65% or the 80% point.

The F1 differences between these high and low vowels (henceforth the high-low F1 difference) were analyzed using a parallel set of statistical tests. As before, the ANOVA assumptions were met (Levene's test, $p > 0.05$, range from 0.147 to 0.771). The main effect of temporal location was significant [$F(4, 132) = 17.6, p < 0.001, \eta^2_p = 0.351$] indicating that the magnitude of the high-low F1 difference was again variable across temporal points, but the descending tendency began from the second temporal point (the 35% point) and not from the first (the 20% point). The main effect of



Color version available online

Fig. 2. Average difference between the high vowel /u/ and the lowest monophthongal vowel (/ɑ/ for either WI or OH variety; /aɪ/, produced as a monophthong, for NC variety) in f0 (**a**) and F1 (**b**). The high-low differences were computed for each individual speaker ($n = 12$ for each regional variety) and for each measurement point in a vowel (20–35–50–65–80%). Each data point represents an average of 36 productions (3 repetitions × 12 speakers).

regional variety was not significant ($p = 0.253$). A significant interaction arose between these 2 factors [$F(8, 132) = 4.8$, $p < 0.001$, $\eta^2_p = 0.224$], which is shown in Figure 2b. An independent *t* test indicated that the high-low F1 difference for NC was significantly greater compared with OH for the first 3 temporal points and was not significant for either the 65% or the 80% point. When compared with WI, the high-low F1 difference for NC was significantly greater only for the first temporal point. The remaining comparisons were not significant.

Summary and Discussion

Experiment 1 established that a large proportion of the total variation in f0 can be predicted by vowel height. This finding supports the physiologically based “intrinsic” aspect of the F1/f0 relationship, in accordance with many previous reports in the literature including those summarized in Whalen and Levitt (1995). However, our “dynamic” exploration of this relationship over the course of a vowel revealed notable differences in the steepness of the regression slope as a function of both temporal measurement location and regional variety, suggesting that the rate of f0 change for each unit of vowel height reflects regional differences in f0 control (i.e., the shape of f0 contour) and in f0 range (i.e., the difference between the intrinsic f0 of high and low vowels) utilized in each regional variety.

We then analyzed the differences between high and low vowels, both in f0 and in F1, to determine whether, for each regional variety, the magnitude of the intrinsic f0 difference corresponds to the magnitude of the F1 difference. Notable discrepancies between the 2 measures were found. In particular, the difference in f0 was smallest for NC and largest for WI, whereas the difference in F1 was largest for NC and smallest for OH, both reaching statistical significance when measured earlier in the vowel, up to the midpoint. The NC case is the most obvious example of a lack of correspondence between the magnitude of the f0 range and that of the F1 range. Mismatches such as this imply that, although the intrinsic aspect of f0 variation is common to all regional

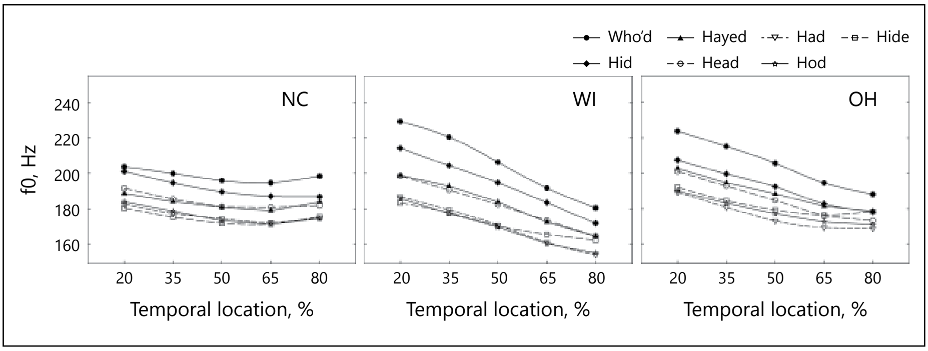


Fig. 3. Average f0 values across 5 temporal points for the set of vowels differing in F1, /u, ɪ, e, ɛ, æ, ɑ, aɪ/ in who'd, hid, hayed, head, had, hod, hide, respectively. The (smoothed) lines connect individual measurement points, approximating f0 shape in each regional variety.

varieties, the magnitude of the intrinsic f0 range is not. A specific f0 range, utilized in a regional speech community, may reflect cultural influences on melodic patterns of a given regional variety spoken in the area. We thus conclude that the magnitude of the f0 range is influenced by regional variation, which confirms our previous finding for these 3 varieties obtained with different high and low vowels (Jacewicz and Fox, 2015a).

Three additional observations warrant further discussion. First, the variation in the steepness of the regression slopes suggests regional differences in f0 shape over the course of a vowel. As an approximation of f0 contours utilized by the 3 varieties, Figure 3 shows average f0 values across all 5 temporal points for several vowels differing in F1, /u, ɪ, e, ɛ, æ, ɑ, aɪ/. We observe that the f0 shape in NC remains relatively flat throughout whereas the pattern in WI shows a substantial f0 fall. The f0 in OH is also falling over the course of a vowel although the fall is comparatively smaller. Thus, regional variation in f0 is manifested not only in differential f0 ranges between high and low vowels, but also in f0 shape throughout the vowel. We underscore that no specific instructions were given to the participants as to how the experimental tokens were to be produced and, given the distinct regional patterns, we infer that these patterns do reflect regional differences.

Second, we observe in Figure 3 that the 3 low vowels /æ, ɑ, aɪ/ tend to overlap in each variety, suggesting that f0 differences diminish among low vowels, irrespective of any differences in their F1. This finding is not surprising and is in line with earlier reports that intrinsic f0 differences in vowels in tone languages are minimized (or "disappear") in the lower part of the f0 range (Connell, 2002; Whalen and Levitt, 1995). However, Figure 3 also shows a similar overlap of the mid vowels /e, ɛ/, which is not only common to all 3 varieties, but again does not reflect the substantial regional variation in F1 of these vowels (Fig. 1). We do not speculate on this curious finding in the absence of any articulatory data, but it is clear that f0 differences can also be minimized in other parts of the f0 range, and not necessarily in the low f0 region. An additional observation is that the magnitude of the regional f0 range seems to play a role in the internal grouping of vowels with respect to their f0 so that the distances between vowels classified as high, mid, and low are enhanced in WI (having the greatest f0 range) and are more "compressed" in NC (having the smallest f0 range).

Finally, we note that the statistically significant regional differences in f0 and F1 were found only for measurements obtained early in a vowel, up to the midpoint (Fig. 2). The F1 drop at the 80% point reflects the effect of a rapid F1 transition to the stop /d/ for low vowels, which reduces the F1 difference between high and low vowels. This is a known effect (Olive et al., 1993), which we have also found in studying a reduction of vowel spaces in these 3 regional varieties as a function of measurement location in a vowel (Fox and Jacewicz, 2008). However, the f0 drop past the vowel's midpoint (at least for WI and OH) is unrelated to the descending F1 transition into /d/ for low vowels, which again results in a mismatch between the "raising" of the later part of the vowel and the "falling" of its f0. This observation provides an additional piece of evidence for a relative independence of f0 control from vowel height. We will return to these issues in the General Discussion, addressing the question of whether vowel f0 can be influenced by regional vowel shifts. We now turn to experiment 2 in order to determine whether regional f0 patterns persist in connected speech.

Experiment 2: f0 Variation in Vowels in Connected Speech

The overall goal of experiment 2 was to determine whether regional varieties differ in the way they implement f0 variation in a vocalic segment in connected speech. To control for this variation, each vowel was produced in 2 different contexts of prosodic prominence, carrying either the main sentence stress or being unstressed. The segmental (microprosodic) influences were controlled for by varying obstruent voicing in the postvocalic position. An extensive set of analyses precluded consideration of all vowel categories used in experiment 1, and only a subset of those vowels was studied in experiment 2.

Speech Materials

Natural meaningful sentences with test words containing 1 of the 5 vowels /ɪ, ɛ, e, æ, aɪ/ were used as stimuli. This set of vowels represents 3 levels of vowel height, and includes tense and lax vowels, monophthongs, and a diphthong (which was produced as such in 2 regional varieties, WI and OH; the variant in NC was produced as a monophthong). As already pointed out, these speech materials were recorded for a related project, and we utilized those samples in the current study. The vowel occurred as a syllable nucleus following a [b] onset and before either a voiced or a voiceless alveolar coda ([dz] or [ts]) in the monosyllabic words *bids, beds, bades, bads, bides* and *bits, bets, baits, bats, bites*, respectively. The words *bit, bait, bet, bat, bite, bid,* and *bed* were all nouns. Unfortunately, *bade, bide,* and *bad* were not, which made it more challenging to create sentences that had exactly the same basic syntactic form. We thus considered these 3 words as nonsense nouns that were defined to speakers prior to their use.

There were 2 types of sentence pairs, in which (1) the target word carried the main sentence stress and the prominent vowel that constituted "the peak of the intonational contour" (Lehiste and Peterson, 1961) was produced with nuclear pitch accent (e.g., *Doc said the small BIRDS are fast. No! Doc said the small BATS are fast*), and (2) the target word occurred in unstressed position and the vowel was unaccented (e.g., *Doc said the small bats are SLOW. No! Doc said the small bats are FAST*). We used this contrastive stress paradigm (Jacewicz et al., 2011b; Lindblom et al., 2007) to elicit the

2 types of vowel production. To make the task more transparent to the participants, the words that carried the main sentence stress were always in all capitals. We emphasize that the phonetic environment was strictly controlled in all sentences: the target vowel was always preceded by a [b] onset and a voiced or voiceless coda (which also marked the plural of the target word), and was always followed by the word "are."

Procedure

The sentence task was a part of a larger recording session and followed immediately the isolated tokens task described in experiment 1. The sentence pairs were presented on a computer monitor in random order. Sentence pairs rather than single sentences were used because our earlier pilot study with untrained noncollege age participants showed that some of them tended to read single sentences with pauses and hesitations. For that reason, meaningful sentence pairs representing a form of minidialogs were created so that the second sentence was semantically related to the first. Given that the second sentence in a pair was typically read with greater fluency, only the target words in the second sentence were selected for acoustic analysis. That is, the stressed and accented vowel was obtained from the sentence such as "No! Doc said the small <u>BATS</u> are fast" and the unstressed and unaccented vowel from "No! Doc said the small <u>bats</u> are FAST" (these target words did not appear with underlines in the stimulus set).

A short practice run was administered prior to the actual test using different sentence pairs. During the practice, the experimenter explained that the word in all capitals was to be emphasized and that the utterances were to be read as naturally as possible ("The second sentence was supposed to be like you would contradict someone"). In order to make the participants more comfortable with the task, they were told they could always repeat what they read if something went wrong or if they were not happy with their production. They were also instructed that they should read fluently and with intended falling intonation (demonstrated by the experimenter). After the practice, the experimenter proceeded with the test sentences. Water was provided to help maintain voice quality as constant as possible throughout the task. Breaks were allowed upon request or when the experimenter noticed symptoms of vocal fatigue. The experimenter monitored each utterance and accepted only fluent productions with appropriate stress placement. The speaker was asked to repeat an utterance in case of disfluencies, incorrect stress placement or nonfalling intonation pattern.

Sixty sentences (representing the second sentence in a pair) were obtained from each speaker except for one who produced half of the set. Out of the 60 sentences, 40 were produced with the target word stressed (5 vowels × 2 consonantal contexts × 4 repetitions) and 20 with the target word unstressed (2 repetitions instead of 4). From the whole data set, 71 target words (3.4%) were excluded due to creaky voice throughout the vowel, mostly in unstressed positions (44 words). Altogether, a total of 2,059 tokens from 36 speakers were analyzed acoustically.

Acoustic Analysis

The digitized tokens were first downsampled to 11.025 kHz and low-pass filtered at 1 kHz. All f0 measures in experiment 2 were based on f0 tracks over the course of the vowel. The f0 tracks were computed using autocorrelation implemented in MATLAB in a series of 16-ms-wide measurement windows with 50% overlap. Vowel onsets and offsets were located by hand and defined using standard segmentation criteria. Several custom MATLAB routines were written to enable a set of f0 analyses and recheck accuracy of f0 measurement. Each resulting f0 contour was examined for any mistracked f0 values that were then hand-corrected using the TF32 and PitchWorks programs. The questionable f0 tracks were recalculated using a separate program written in MATLAB. A final reliability check was done for all f0 tracks on all (100%) experimental tokens by the second author. Any corrections were made prior to further data processing in spreadsheets.

The acoustic variables included vowel duration, overall mean f0 and a set of measures characterizing f0 shape. These measures were of 2 types. First, landmark measures of f0 shape included the initial f0 at vowel onset, peak f0 and the final f0 at vowel offset. From these f0 landmarks, a set of dynamic measures of f0 shape were derived. The use of running speech in experiment 2 necessitated a more detailed analysis of entire f0 tracks rather than sampling f0 at selected temporal points as in experiment 1, an approach suitable for vowels produced in isolated tokens.

Landmark Measures of f0 Shape

The overall mean f0 (henceforth overall f0) was used as a measure of the speaker's mean f0 for the target vowels. This was simply the average f0 across all measurement windows for each individual vowel. Given that the mean is calculated from f0 values over the vowel's entire duration, stressed vowels with the greater f0 excursion for the pitch accent are expected to have higher f0 than unstressed vowels, whose f0 contours are typically flat. The initial f0 value, onset f0 (in hertz), was that obtained at the first measurement window following vowel onset. We decided to use the onset f0 measure for both stressed and unstressed vowels to test for possible regional effects also in unstressed vowels, whose flat f0 contours were not expected to provide additional insights when sampled at later time points in a vowel. Peak f0 (or f0 maximum) was the highest detected f0 value in a stressed vowel. The offset f0 was the f0 value obtained for the last measurement window in a stressed vowel.

Dynamic Measures of f0 Shape

A set of dynamic measures of f0 shape was further derived from the acoustic landmarks. f0 rise was the f0 change from onset to peak, and f0 fall was the f0 change from peak to offset. These 2 measures were expressed both in hertz and in cents on a semitone scale (details to be provided below). The temporal location of the peak (representing the first occurrence of the highest detected f0) was calculated in 2 ways, in terms of (1) the absolute time (in milliseconds) from vowel onset and (2) its relative position with regard to the duration of the vowel (with values ranging from 0 to 100%).

Statistical Analysis

A repeated-measures ANOVA was used to assess the statistical significance of the within-subject factors vowel, coda voicing (voiced, voiceless) and stress (if applicable) and the between-subject factor regional variety. Specific analyses will be described in the sections below. The overall design was balanced and fully crossed, and therefore no transformation of dependent variables was considered in those cases in which normal distribution of residuals was moderately violated. The homogeneity of variance was tested with Levene's test; out of 120 groups, error variance of the dependent variable was equal for 107 ($p > 0.05$) and unequal for only 13 ($p < 0.05$). To

Table 2. Summary of significant main effects and interactions from repeated-measures ANOVAs for vowel duration, overall mean f0 and onset f0

Main effects and interactions	Vowel duration	Overall mean f0	Onset f0
Stress	0.784***	0.847***	0.788***
Vowel	0.943***	0.520***	0.596***
Coda voicing	0.897***	–	–
Variety	0.213*	–	–
Stress × coda	0.795***	0.433***	0.607***
Stress × vowel	0.324***	0.434***	0.473***
Vowel × coda	0.386***	–	–
Stress × coda × vowel	0.114**	–	–
Stress × variety	–	–	–
Coda × variety	–	0.369***	–
Vowel × variety	0.326***	0.160**	0.170**
Stress × coda × variety	–	–	–
Stress × vowel × variety	0.168**	–	–
Coda × vowel × variety	0.179**	0.195***	0.139**
Stress × coda × vowel × variety	0.112*	–	–

The analyses include both stressed and unstressed vowels. Shown are partial η^2 values. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

protect against type I error rate inflation due to the multiple analyses of f0 and small homogeneity violations, a more stringent significance level was adopted to reject the null hypothesis, and significance was assumed at an alpha level of 0.01. However, since no multiple analyses were conducted for vowel duration as a different dependent variable, the more common alpha level of 0.05 was assumed for vowel duration. For the 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 (Mauchly's test). Unless otherwise stated, all of the post hoc *t* tests used the Bonferroni adjustment for multiple comparisons. All analyses were performed using IBM SPSS Statistics v. 24. In addition to indicating *p* values for specific *F* tests, partial η^2 values are provided for all significant main and interaction effects (Pierce et al., 2004).

Results

Vowel Duration

We begin with vowel duration, which represents the time span over which f0 measurements were made. The ANOVA results are summarized in Table 2. As expected, all main effects were significant. Stressed vowels were significantly longer than unstressed vowels (mean = 207 vs. mean = 156 ms). Vowels before voiced consonants were significantly longer than before voiceless consonants (mean = 205 vs. mean = 158 ms). A significant main effect of vowel exposed the well-known inverse relationship between vowel duration and vowel height (or degree of openness): high vowels were shorter, and low vowels were comparatively longer. The main effect of variety indicated that the NC vowels (mean = 204 ms) were significantly longer than either OH (mean = 170 ms) or WI variants (mean = 171 ms), but the latter 2 did not differ significantly from

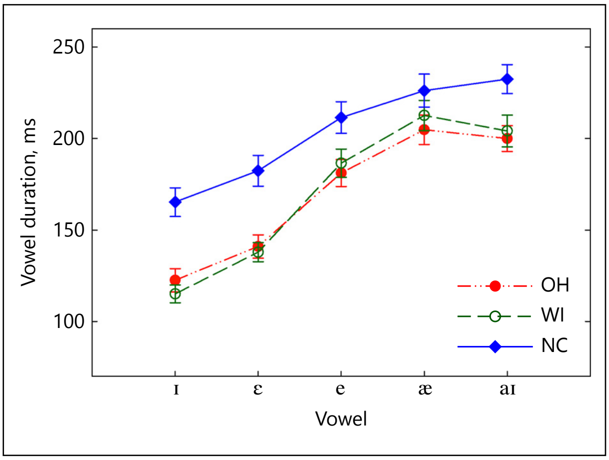


Fig. 4. Average (±SE) duration of vowels produced by Ohio (OH), Wisconsin (WI), and North Carolina (NC) speakers.

one another (Scheffé test). This finding is consistent with our previous reports on these 3 varieties (Jacewicz et al., 2007). As illustrated in Figure 4, the differences between the NC and the 2 other varieties tended to be greater for short vowels /ɪ, ɛ/ than for long vowels (except for /aɪ/), which was the locus of a significant vowel by variety interaction. This result is also consistent with our previous finding with different participants from these 3 varieties (Jacewicz and Fox, 2015c).

Several remaining significant 2-, 3-, and 4-way interactions reported in Table 2 are not of immediate interest. We only note that some of these interactions can be predicted, for example that duration differences due to coda voicing will be greater when the vowels are stressed and smaller when they are unstressed (stress × coda voicing), and that these differences will be greater for long vowels than for short vowels (vowel × coda voicing). We can also expect that the magnitude of the difference due to stress will be somewhat variable for individual vowels (stress × vowel), and that it will further interact with coda voicing (stress × coda voicing × vowel) and regional variety, given that NC vowels were not only comparatively longer but that the differences between short and long vowels were smaller in NC than in the 2 other varieties (stress × coda voicing × vowel × variety).

Overall f0 and Onset f0

Turning to f0 measures, we first consider the results of separate ANOVAs for overall f0 and onset f0 (Table 2). Notably, all significant main effects and all but 1 significant interactions were consistent for both measures. As expected, stress had a strong effect on f0: stressed vowels had a significantly higher mean f0 than unstressed vowels. There was also a significant main effect of vowel that reflected differences in intrinsic f0 in the expected direction: high vowels had a higher f0 than mid and low vowels, respectively. These differences were further explored with a paired t test. For overall f0, all but 3 differences were significant at the Bonferroni-adjusted level of 0.001. These *nonsignificant* differences were between the 2 mid vowels /ɛ, e/ ($p = 0.061$), the 2 low vowels /æ, aɪ/ ($p = 0.454$) and for the /e, aɪ/ pair ($p = 0.011$). For onset f0, only the differences between /ɛ, e/ ($p = 0.023$) and /æ, aɪ/ ($p = 0.619$) were not significant. All remaining comparisons were significant. This overall pattern suggests

DOI: 10.1159/000484610

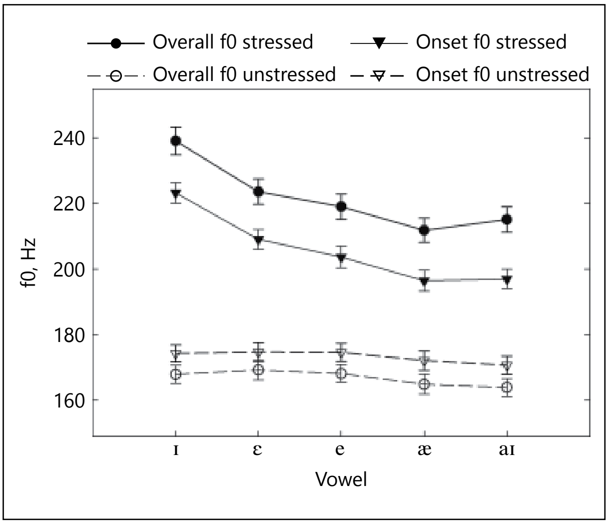


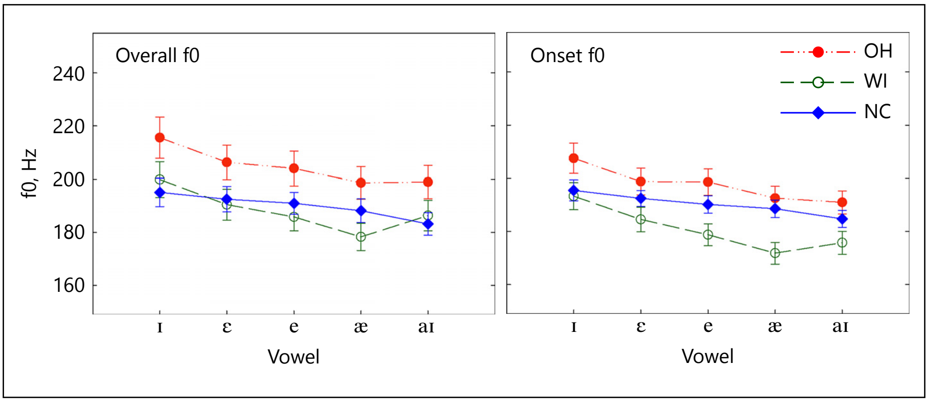
Fig. 5. Average (±SE) overall f0 and onset f0 in vowels as a function of stress.

that f0 differences diminish not only among low vowels (Ladd and Silverman, 1984; Whalen and Levitt, 1995), but also among mid vowels, consistent with the finding in experiment 1 (Fig. 3).

However, a significant stress × vowel interaction, illustrated in Figure 5, clarified that the intrinsic f0 variations associated with differences in vowel height were present only in stressed vowels. As is also evident in Figure 5, f0 differences between stressed and unstressed variants of each vowel were greater for overall f0 than for onset f0, although the overall pattern was common to both.

The main effect of coda voicing was not significant for either f0 measure. However, a significant stress × coda interaction revealed that a voiceless coda enhanced the f0 difference between stressed and unstressed vowels. In particular, before a voiceless coda, f0 was higher in stressed vowels and lower in unstressed vowels, producing a difference of 59 Hz for overall f0 and 37 Hz for onset f0. A voiced coda comparatively decreased the f0 difference between stressed and unstressed vowels to 51 Hz for overall f0 and 28 Hz for onset f0.

A significant coda × variety interaction was found only for overall f0. This interaction showed that f0 was higher before a voiceless coda for both OH and WI but not for NC, where f0 was higher before a voiced coda. Furthermore, this general pattern was more variable for individual vowel categories, which was the locus of a significant 3-way coda voicing × vowel × variety interaction, also found for onset f0. This interaction revealed mixed patterns. For overall f0, all OH vowels had a higher f0 before a voiceless coda and all NC vowels had a *lower* (not higher) f0 before a voiceless coda. The effects were mixed and vowel dependent for WI. Measured at the onset, f0 was higher before a voiceless coda for 4 OH vowels, 3 WI vowels, and 1 NC vowel, and for the remaining vowels, the effects of coda voicing were either minimized or f0 was higher before a voiced coda. We underscore the small effect size of the interaction for the onset f0 measure, which may reflect some "residual" effects rather than a more systematic variation. The effects of coda voicing will be explored in greater detail in a set of separate analyses of f0 shape.



Color version available online

Fig. 6. Average (±SE) overall f0 and onset f0 in vowels as a function of regional variety.

Finally, the main effect of variety was not significant. However, important f0 differences showed up in a significant vowel × variety interaction, both for overall f0 and onset f0, which is visualized in Figure 6. Exploring this interaction, paired *t* tests revealed the following patterns for each regional variety. For WI, significant differences for overall f0 were between /ɪ/ and each of the remaining vowels; also significant was the difference between mid and low vowels /ɛ, æ/. The remaining comparisons were not significant at the Bonferroni-adjusted level of 0.001. However, for the f0 onset measure, the overall pattern was like that for the main effect of vowel: nonsignificant differences were found only between the 2 mid vowels /ɛ, e/ ($p = 0.006$), the 2 low vowels /æ, aɪ/ ($p = 0.046$), and for the /e, aɪ/ pair ($p = 0.124$). For OH, significant differences for overall f0 were between /ɪ, e/, /ɪ, æ/, and /ɪ, aɪ/. At the onset, only the differences between high and low vowels were significant, /ɪ, æ/, and /ɪ, aɪ/. The remaining comparisons were not significant. A very different pattern emerged for NC in that *none* of the pairwise comparisons were significant for either f0 measure.

Overall, the vowel × variety interaction revealed that, although the general trend in intrinsic f0 related to the vowel's height was maintained in all 3 varieties, regional variation was manifested in the magnitude of f0 differences between high, mid, and low vowels. In particular, based on their f0 (especially onset f0), WI vowels could be generally separated into 3 broad categories of height – high, mid, and low – whereas OH vowels exhibited mostly 2 levels of height – high and low. f0 differences among NC vowels were too small to afford any height distinctions, and they could be regarded as representing 1 level of height. There is a good correspondence between these broad levels of height and the magnitude of the f0 difference between high and low vowels. In particular, the f0 differences between /ɪ, æ/ were greatest for WI (22 Hz for overall f0 and 21 Hz for onset f0), were comparatively reduced for OH (17 Hz for overall f0 and 15 Hz for onset f0), and were minimal for NC (7 Hz for each measure). Clearly, this overall pattern of intrinsic f0 differences as a function of regional variety, although reduced in magnitude, corresponds to that in isolated words in experiment 1. We need to bear in mind, however, that both stressed and unstressed vowels were included in the above computations so that f0 values may also reflect regional variation in f0 shape. We expect the analyses of f0 shape below to provide more insights into this issue.

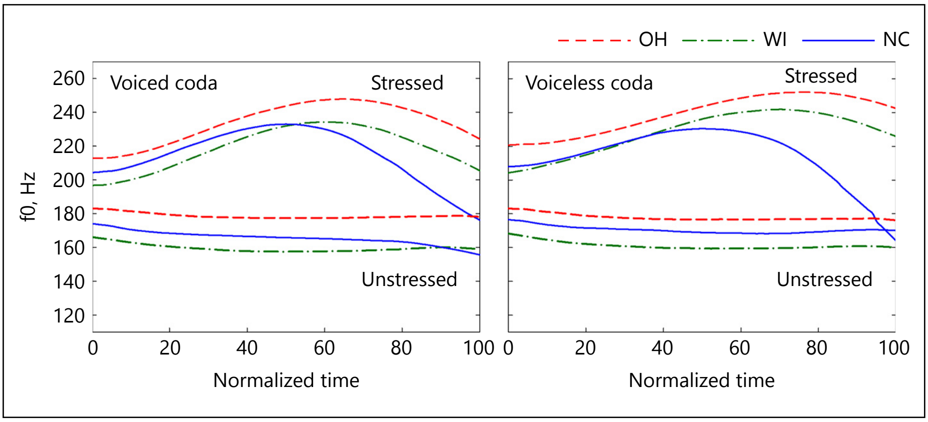


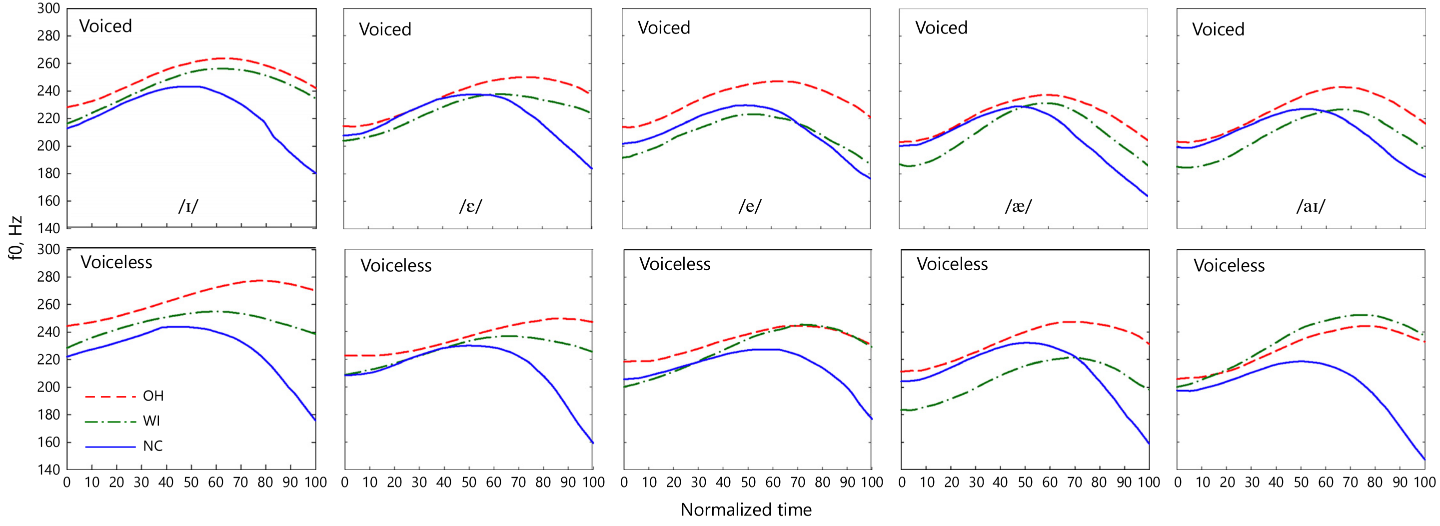
Fig. 7. Average smoothed time-normalized f0 contours collapsed for all stressed (*n* = 1,403) and all unstressed (*n* = 656) vowels /ɪ, ɛ, e, æ, aɪ/ in the context of voiced and voiceless obstruent consonants in the coda. The contours were generated from all measurements time-normalized to a 0- to 100-point scale based on time proportions for each vowel (with values between actual measurements linearly interpolated).

In summary, stress had the strongest effect on f0, which was manifested in greatest effect sizes for both overall f0 and onset f0 measures (Table 2). Also strong was the effect of vowel height. However, the effects of coda voicing and regional variety were manifested in interactions rather than in main effects, suggesting that any differences due to these factors may be more subtle. For example, they may occur at later portions of the vowel rather than at vowel onset and may be obscured by averaging the f0 over the vowel's duration. A curious finding was that overall f0 was significantly higher before voiceless coda but this was true only for 2 varieties, OH and WI. Unexpectedly, the opposite was found for NC, where the overall f0 was higher before a voiced coda. To better understand the sources of the complex interactions between vowel f0 and regional variety in the context of coda voicing, we now turn to the results for the extended set of measures of f0 shape.

The Landmarks of f0 Shape

Figure 7 shows time-normalized f0 contours averaged across speakers. These time-normalized contours, created for the purpose of making detailed graphical comparisons, were generated using 101 time points (from 0 to 100), linear interpolation, and smoothing. Time normalization allows the averaging of data points (f0 contours) across repetitions and speakers, which reduces random variations unintended by the speaker and leaves only consistent tonal variations (Xu, 1997, 2015). The specific measurements of f0 shape used in the statistical analyses were of course obtained from non-time-normalized contours.

Both stressed and unstressed time-normalized contours are displayed in Figure 7 to explicate variation in f0 as a function of stress and how this variation is manifested in all 3 varieties. As is evident, the f0 contours of stressed vowels reflect changes in pitch over time but those in unstressed vowels remain relatively flat throughout in all 3 varieties. Given the relative constancy of f0 and statistically nonsignificant effects



Color version available online

Fig. 8. Average smoothed time-normalized f0 contours for individual stressed vowels in the context of a voiced and a voiceless coda for Ohio (OH), Wisconsin (WI), and North Carolina (NC) speakers.

of the factors of interest, we decided to exclude the unstressed variants from further consideration in this paper and to explore only the f0 contours in stressed vowels. In so doing, we now consider the landmark measures of f0 shape at peak f0 and offset f0. The data were analyzed using 3-way ANOVAs that excluded stress as a within-subject factor.

Peak f0. Figure 8 displays the time-normalized average f0 contours of individual vowels in the 3 varieties. An ANOVA for peak f0 returned a significant main effect of vowel [$F(4, 132) = 25.4$, $p < 0.001$, $\eta^2_p = 0.435$], showing that the f0 maxima for the 2 short vowels /ɪ, ɛ/ were significantly higher than for the remaining longer vowels /e, æ, aɪ/. However, while peak f0 for /ɪ/ also differed significantly from /ɛ/, none of the pairwise differences were significant for the longer vowels. There was also a significant 3-way interaction between vowel, coda voicing, and variety [$F(8, 132) = 4.4$, $p < 0.001$, $\eta^2_p = 0.209$]. This interaction did not yield any consistent pattern, however.

Offset f0. f0 variability was far greater at the offset than at the peak (compare f0 values in Fig. 8). The main effect of vowel was significant [$F(4, 132) = 24.1$, $p < 0.001$, $\eta^2_p = 0.422$] and indicated that the pattern of intrinsic f0 variations in vowels was generally maintained also at the offset. Importantly, a significant main effect of variety [$F(2, 33) = 6.8$, $p = 0.003$, $\eta^2_p = 0.291$] showed that NC f0 offsets were significantly and substantially lower (mean = 170 Hz) than either OH (mean = 229 Hz) or WI (mean = 212 Hz); the latter 2 did not differ from each other on the Scheffé test. The robust f0 drop in NC vowels demonstrated that f0 contours in this variety were markedly different. Of note is a significant coda voicing × variety interaction [$F(2, 33) = 7.4$, $p = 0.002$, $\eta^2_p = 0.311$] that arose because f0 before a voiceless coda was higher than before a voiced coda for OH and WI and this pattern was reversed for NC. This interaction revealed that regional varieties can differ in the effects of coda voicing. In particular, the f0 at the offset can be raised before voiceless obstruents in some varieties (OH, WI) and can be lowered in the same context in others (NC). The 2 remaining interactions were also significant but they did not produce additional insights and are not discussed.

The results of the landmark analyses for the 2 f0 measures, peak f0 and offset f0, revealed that the differences between NC and the 2 other varieties were particularly exaggerated at the offsets, which also demonstrated differential effects of coda voicing.

Variation in the Temporal Location of Peak f0

The observed regional discrepancies with regard to the effects of coda voicing are difficult to interpret without consideration of the entire time course of the f0 contour. It appears that analyzing f0 variation at selected landmark points is incomplete and may lead to unnecessary confusions and misinterpretations. Our final set of analyses seeks to characterize the dynamics of f0 contour shape to better understand these f0 patterns.

We first consider variation in the temporal location of peak f0 to observe its alignment across the 3 varieties. We expect the location of the peak to be variable in the current design due to the effects of the intrinsic vowel duration and coda voicing. The existing literature guides us in our expectations that, if the peak is timed with respect to the entire syllable, it will be moved closer to vowel onset for intrinsically short vowels and closer to the offset for long vowels and diphthongs (e.g., House and Wichmann, 1996; Ladd, 1996). In other words, peak f0 will appear earlier in time in short vowels (having a shorter rise time) and will be delayed in long vowels (having a longer rise

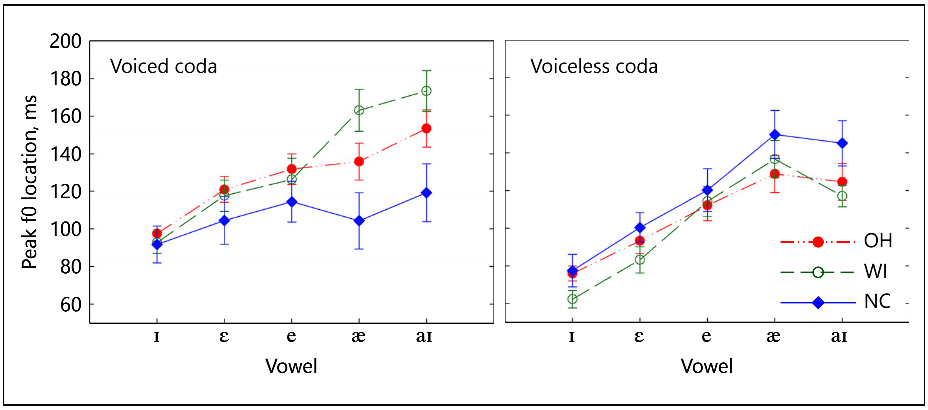


Fig. 9. Average (±SE) absolute temporal peak f0 location for stressed vowels (representing f0 rise time) in the context of a voiced and a voiceless coda for Ohio (OH), Wisconsin (WI), and North Carolina (NC) speakers.

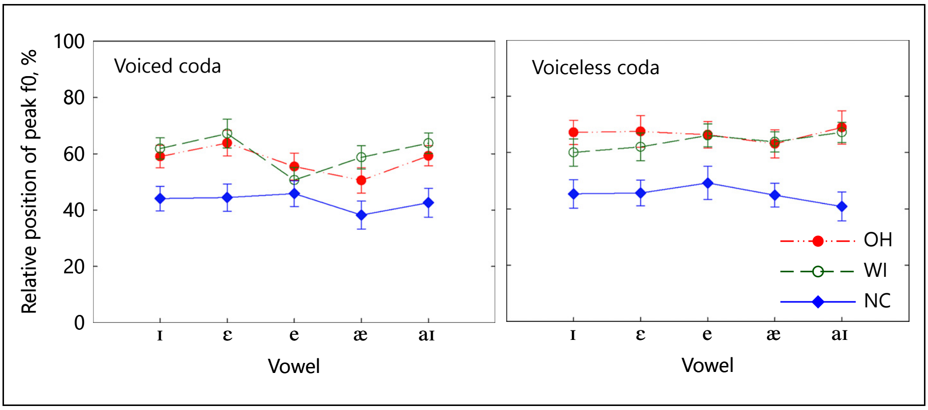
time). We further expect that a voiceless coda will also push the peak earlier as the vowels are shortened before voiceless consonants.

We first analyzed these temporal variations measured in absolute time (in milliseconds). The results are displayed in Figure 9. As expected, the peaks appeared earlier in the short vowels /ɪ, ɛ/ and were delayed in the longer vowels /e, æ, aɪ/, thus vowel duration and temporal location of the peak seemed positively covarying. However, this covariance was somewhat distorted by regional variation. The temporal location of the peak also varied predictably with coda voicing, at least for OH and WI. However, in NC vowels the peaks were aligned earlier (and not later) before a voiced coda than before a voiceless coda.

These mixed results with regard to the effects of regional variety prompted us to re-examine the temporal location of the peak in terms of its relative – rather than absolute – position in a vowel. The relative position of peak f0 with respect to each vowel's duration (in percent) is informative because it eliminates the temporal variation related to inherent vowel length, coda voicing and regional variety, and possible effects of speaking rate differences across participants.

As shown in Figure 10, most differences in peak location related to vowel length were eliminated in this analysis. The main effect of vowel was not significant as the peaks were on average located between 53 and 58% in a vowel. The effect of coda voicing was significant [$F(1, 33) = 19.32$, $p < 0.001$, $\eta^2_p = 0.369$], showing that the peaks occurred closer to vowel offset before a voiceless coda (mean = 59%) than a voiced coda (mean = 54%). The main effect of variety was significant [$F(1, 33) = 7.10$, $p < 0.001$, $\eta^2_p = 0.301$] and the Scheffé test revealed that NC peaks occurred earlier (mean = 44%) than either OH or WI, which did not differ significantly from one another (both mean = 62%).

In summary, the analysis of the relative position of the peak yielded different results than when the position of the peak was measured in absolute time. For all vowels, the peaks were aligned in a central portion of the vowel, shortly after vowel midpoint, regardless of whether the vowels were short or long. The peaks were also aligned



Color version available online

Fig. 10. Average (±SE) relative position of peak f0 within a vowel (in percent) for individual stressed vowels in the context of a voiced and a voiceless coda for Ohio (OH), Wisconsin (WI), and North Carolina (NC) speakers.

later and not *earlier* before a voiceless than a voiced coda. Most importantly, there were robust differences related to regional variety in that NC peaks occurred earlier than the peaks in the other varieties, reflecting the comparatively longer fall time from peak to offset f0 in NC vowels.

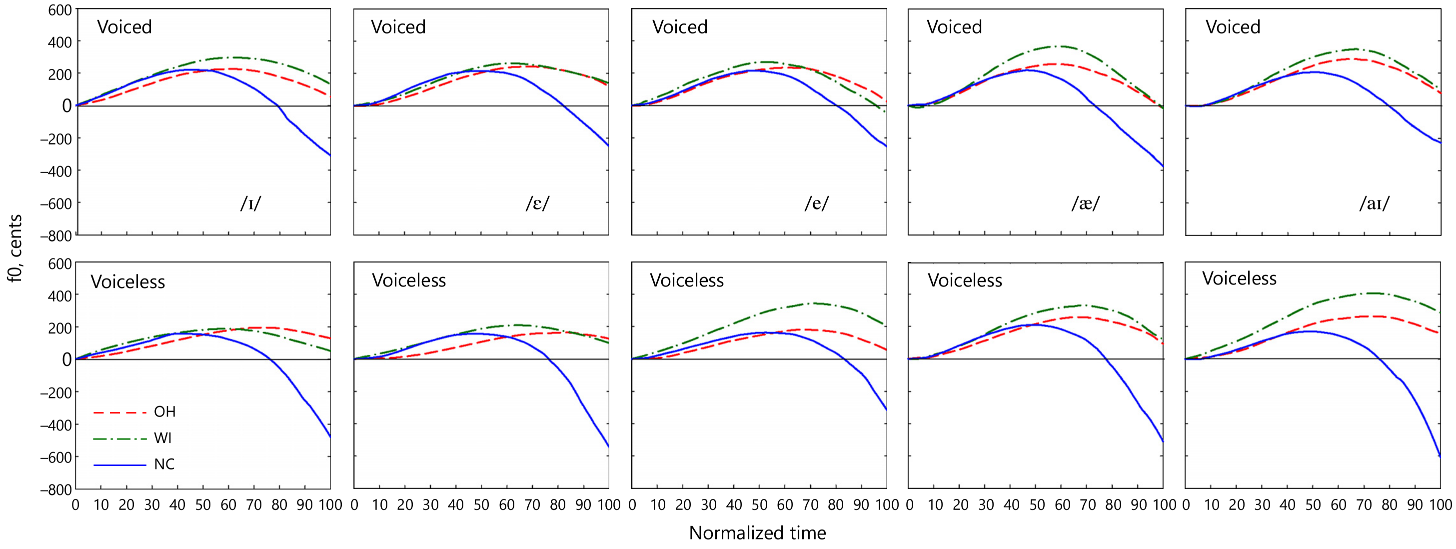
The Dynamics of f0 Shape

We now consider the amount of f0 change for the rise from onset to peak (f0 rise) and f0 fall from peak to offset (f0 fall). All analyses were initially done using the original f0 values in hertz. However, given differences in basic f0 among speakers related to physiological features as well as additional variation in f0 related to vowel height and regional variation, the f0 changes were re-analyzed in cents on the semitone scale as a type of speaker-intrinsic, vowel-intrinsic normalization for this extraneous variation. The musical semitone scale (which is a logarithmic transformation of the hertz scale) has proven successful in assessing productions of intonationally equivalent utterances (e.g., Nolan, 2003) as it relates to psychoacoustics of speech. We converted the f0 contour in hertz to f0 contour in semitones (cents) relative to the f0 onset frequency (i.e., the f0 in the first measurement window) in each vowel using equation 1:

$$\mathrm{f0}_{\mathrm{cents}} = 1{,}200 \times \log_2 (\mathrm{f0}_{\mathrm{n}}/\mathrm{f0}_{\mathrm{onset}}) \qquad (1)$$

where $\mathrm{f0}_{\mathrm{onset}}$ is frequency (in hertz) of the onset f0 and $\mathrm{f0}_{\mathrm{n}}$ is the frequency of f0 (in hertz) in the measurement interval being converted. This represents a measure of f0 change relative to the f0 onset value. Figure 11 shows the f0 contours from Figure 8, rescaled from hertz to semitones (cents).

f0 Rise. A 3-way ANOVA for f0 rise (in semitones) returned only a significant main effect of vowel [$F(4, 132) = 8.3$, $p < 0.001$, $\eta^2_p = 0.195$]. The amount of f0 rise was inversely related to vowel height so that the rise was smallest in high vowels and greatest in low vowels. The mean f0 rise increased progressively for /ɪ, ɛ, e, æ, aɪ/, in that order. Regional variations in the size of f0 rise were small and not significant.



Color version available online

Fig. 11. Average smoothed time-normalized f0 contours measured in semitones (cents) for individual stressed vowels in the context of a voiced and a voiceless coda for Ohio (OH), Wisconsin (WI), and North Carolina (NC) speakers.

f0 Fall. An ANOVA for f0 fall returned a significant main effect of vowel [$F(3.2, 106.9) = 5.8$, $p = 0.001$, $\eta^2_p = 0.148$]. The f0 drop was significantly greater for the vowels /æ, aɪ/ than for the remaining vowels, which did not differ significantly from one another. A significant main effect of variety [$F(2, 33) = 20.1$, $p < 0.001$, $\eta^2_p = 0.549$] revealed a major difference in that NC vowels had a significantly greater f0 fall than either WI or OH, which did not differ significantly from one another (Scheffé test). A significant coda voicing × variety interaction [$F(2, 33) = 6.2$, $p = 0.005$, $\eta^2_p = 0.273$] further revealed that the f0 fall in NC vowels was greater before a voiceless than a voiced coda while the reverse was true for both OH and WI.

These discrepancies can be explained on the basis of variation in f0 fall time. In particular, given that the f0 fall in both OH and WI occurred in about half the time as in NC, an additional shortening before a voiceless coda caused the OH and WI contours to be "cut short" (Hombert et al., 1979), which resulted in a higher terminal f0 and a smaller f0 drop. Conversely, an increased duration before a voiced coda allowed more time for a comparatively greater f0 drop. However, given that NC f0 fall time was almost twice as long, the shortening before a voiceless coda did not pose constraints on completion of the full contour and did not cause "clipping" of the terminal frequency as in the 2 other varieties. Rather, it appears that the shortened duration before a voiceless coda contributed to a greater f0 change from peak to offset. Likewise, a longer duration of f0 fall before a voiced coda permitted more time for the completion of the contour, producing a comparatively smaller f0 drop. Clearly, the results for the f0 fall provide strong evidence for the robust difference in the way regional varieties complete the f0 contours under time constraints: NC speakers produce the full contour even when the duration of a vowel is shortened before a voiceless consonant whereas both OH and WI speakers truncate their contours in identical contexts.

General Discussion

In the current paper, we tested the general hypothesis that f0 of a vocalic segment varies systematically as a function of regional variation in American English. This hypothesis was tested using an extended set of vowels produced in isolated hVd tokens (experiment 1) and a selected subset of vowels in connected speech (experiment 2). The following major findings support our hypothesis.

First, regional differences were found in isolated tokens. The differences were in pitch patterns over the course of a vowel in that f0 shape in NC was relatively flat whereas both WI and OH exhibited a falling f0, with the f0 fall in WI being comparatively greater. In addition, regional varieties were found in the use a specific f0 range (measured as the difference between the intrinsic f0 of high and low vowels), whose magnitude did not correspond to the magnitude of the F1 difference between these vowels. For example, the difference in f0 could be small but the difference in F1 could be large as in NC, or the difference in f0 could be large and the difference in F1 could be smaller as in WI. As a whole, the results of experiment 1 suggest regional differences in f0 control for vowels produced in isolated syllables or words.

Second, the f0 patterns produced in isolated tokens were not preserved in connected speech. Rather, f0 shapes in vowels carrying the main sentence stress were different in nature so that NC speakers produced full contours with a very large f0 drop whereas the contours in WI and OH speakers were comparatively shallower and appeared "unfinished." Undoubtedly, these differences reflect regional variation in the use of prosody, including f0 peak alignment. However, no significant differences due to regional variation were found when the vowels were unstressed as their f0 contours remained flat over the course of vowel durations. Consistent with experiment 1, regional differences in the utilized f0 ranges were also preserved in connected speech.

Third, our exploration of the influence of obstruent consonant voicing on the f0 of the preceding stressed vowels provided new evidence that regional varieties also differ in the implementation of the effects of consonantal context. We admitted 3 possibilities for the effects of a voiceless coda: (1) the f0 contour is "clipped," and terminal f0 values are raised in all dialects; (2) the contour is not "clipped," and the dialect-specific f0 shape that appears before a voiced coda is preserved; (3) some varieties may "clip" before a voiceless coda and some may maintain the same f0 shape in both contexts. The current results presented us with a fourth possibility: some varieties "clip" the f0 contour (and terminal f0 is raised) and some complete the entire f0 shape in a shorter time frame, which produces a larger f0 drop than before a voiced coda (and terminal f0 is lowered).

While the effects of coda voicing were apparent in overall f0 showing that f0 was significantly higher before a voiceless coda in OH and WI, an unexpected pattern emerged for NC, where the overall f0 was higher before a voiced and not voiceless coda. It was the second set of analyses exploring the dynamics of f0 shape that provided us with data critical to the interpretation of these regional variations in f0. In particular, coda voicing did not have a strong influence on the f0 rise but it had a profound effect on f0 drop from peak to offset, and these consonantal context effects interacted with regional differences in f0 use. For both OH and WI speakers, f0 peaks occurred later in time and the shift in peak alignment contributed to a "clipping" of the contour before a voiceless coda so that the falling accent shape before a voiced coda appeared "unfinished." However, a different strategy was used by NC speakers who did produce the complete f0 contour in the shorter amount of time available before a voiceless coda. The f0 drop from peak to offset before a voiceless coda was particularly large, which resulted in much lower terminal f0 values compared with those before a voiced coda. This robust difference explains why the terminal f0 was significantly lower (and not higher) before a voiceless coda for NC vowels. Figure 12 illustrates these regional differences at the vowel's offset, plotted in absolute time. While the magnitude of "clipping" of the contour before a voiceless coda varied with vowel category, the overall pattern for OH and WI was different than that for NC, suggesting that completion of the f0 contour in the context of coda voicing can be influenced by regional variation.

Vowel duration has not received much consideration in this paper due to the primary focus on f0 patterns. We need to point out, however, that the size of temporal variation in vowels as a function of either stress or coda voicing was comparable across the 3 varieties despite the fact that NC vowels were inherently longer than either OH or WI. In particular, the proportional relations between durations of stressed and unstressed vowels were common to all so that stressed vowels were on average 1.3 times greater than unstressed vowels. This result was reflected in a nonsignificant interaction between stress and variety ($p = 0.815$). Similarly, durations before voiced coda were also about 1.3 greater than durations before voiceless coda, and the interaction between coda and variety was not significant ($p = 0.554$). These common proportional temporal relations underscore the fact that regional differences in f0 shape could

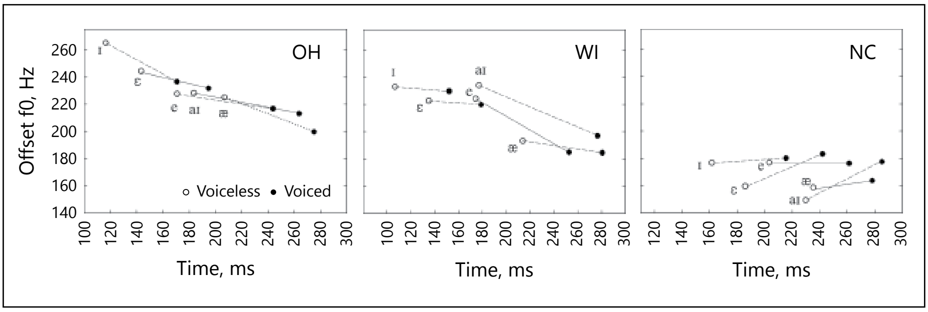


Fig. 12. Average terminal f0 for individual vowels in the context of coda voicing plotted in absolute time (milliseconds) for Ohio (OH), Wisconsin (WI), and North Carolina (NC) speakers.

not be influenced by regional influences on temporal variation in vowels. That is, if vowel duration was shortened before a voiceless coda, the amount of the reduction was common to all 3 varieties.

Can Regional Vowel Shifts Influence Vowel f0?

Our exploration of the relationship between F1 and f0 in experiment 1 allows us to formulate realistic predictions as to whether f0 in vowels involved in regional shifts can be influenced by positional changes in vowel height. Based on the results, it cannot be assumed that raising or lowering of a vowel (i.e., a change in F1) will automatically involve a corresponding change in f0. Rather, different regional varieties may utilize different f0 ranges between high and low vowels, and the magnitudes of these f0 ranges will largely determine the degree of overlap or separation of individual vowels with respect to their intrinsic f0. We also need to bear in mind that any f0 differences will be minimized in the low region of the vowel space (e.g., Ladd and Silverman, 1984) so that even the most raised variant of /æ/ such as in NC or WI will overlap in f0 with other low vowels in each variety, including /ɔ, ɑ/ (see Fig. 1 and 3). Furthermore, mid vowels /e, ɛ/ will also overlap in f0, which implies that any effects of the Southern Shift will not be manifested in f0 of these vowels; nor will the reversal of /i, ɪ/ be reflected in f0 patterns, given that the f0 range is greatly reduced in the NC variety and intrinsic f0 distinctions are smaller than, for example, in the WI variety.

Importantly, to determine the magnitude of a regional f0 range, f0 measurements in high and low vowels need to be obtained early in a vowel, before the midpoint, where regional differences can reach statistical significance. But this temporal location will not be ideal when studying regional variation in f0 in vowels in connected speech, where the effects of prosody and segmental context will largely interact and shift f0 maxima to different temporal positions. It is also recognized that f0 measured at the energy peak of the syllable (i.e., at the rms peak located typically before the midpoint) will unlikely reflect the peak of the f0 contour that is typically reached later in time than the energy peak (Jacewicz and Fox, 2008b; Ladd and Silverman, 1984). Thus, there are no strict rules as to the only ("one size fits all") method for sampling f0 in a vowel. As we have shown in this paper, different f0 measures serve different purposes and individual researchers will need to determine the suitability of particular methodological choices to the objectives of their studies.

Little discussion thus far has been devoted to diphthongs that also mark distinctions among regional varieties. In general, the results suggest that time-varying spectral changes in a vowel are "disjoint" from its f0 shape, indicating that f0 control is separate from the changing configuration of the vocal tract over the course of a vowel. However, we observed that f0 in the later parts of /aɪ, aʊ/ in WI and OH did rise slightly, although never reaching the f0 value of the high vowel target /ɪ, ʊ/. But it is also the case that the terminal f0 values of high vowels become much lower due to a substantial f0 fall, which makes it difficult to determine whether the second high f0 target can ever be reached in a diphthong. This issue awaits separate experimental investigation in the future.

Concluding Remarks, Caveats, and Future Directions

As discussed earlier in the paper, the intrinsic aspect of f0 variation in relation to vowel height was common to all regional varieties. We do not return to this issue here given extensive prior discussions of experiment 1, but one important methodological concern has not been addressed thus far. Namely, the accuracy of formant estimation by linear prediction techniques is prone to error and drawbacks of LPC analysis are well known (e.g., Atal, 1975; Vallabha and Tuller, 2002). In particular, it is recognized that formant frequencies are biased by f0 and that the problem will escalate in the F1 region due to the dominance of the most intense harmonic. There are currently no feasible solutions to this problem. Sociophonetic investigations require large data sets of natural speech, and analyzing such "big data" by manual methods is too labor intense as to ensure steady progress in the field. Besides, as demonstrated by Shadle et al. (2016), hand measurements of F1 by experts also tend to be biased by the frequency of the nearest harmonic of the f0. The newer manual methods, still in early stages of small-scale demonstrations (Alku et al., 2013; Fullop, 2010, 2011), have been shown to improve the accuracy of formant estimation in synthetic vowels and in naturally produced monophthongs in several speakers but, "unfortunately, no automatic means of extracting those values has been found to be reliable" (Shadle et al., 2016, p. 726).

In the absence of reliable and statistically validated automatic tracking techniques, we followed the accepted conventional LPC analysis methods in formant estimation implemented in MATLAB, using hand correction in comparison to FFT spectra, wideband spectrograms (with superimposed LPC formant locations) and a second LPC analysis with filter coefficient adjustment (in TF32), with the consistency of measurement applied on a speaker-by-speaker basis. With regard to F1 and f0, these analyses produced results comparable with those found in the earlier literature (e.g., de Boer, 2011; Hillenbrand et al., 1995; Whalen and Levitt, 1995). Admittedly, although extensive visual inspections and manual manipulations of analysis parameters were applied to reduce error in this study, an error-free formant tracking using LPC is still an unrealistic expectation, and we hope that further progress in the development of more sensitive algorithms will reduce measurement error in the future. However, automatically tracking the formant change over the time course of the vowel using improved linear prediction techniques such as in Alku et al. (2013) and Liu and Shimamura (2015) is likely to remain a challenge for some time.

Even with these limitations, the analysis of the F1/f0 relationship in the current study does provide some degree of confidence that the LPC technique allows for reasonably accurate measurements of the first formant, since the accuracy begins to deteriorate seriously with higher f0 values, about 400 Hz (Monsen and Engebretson, 1983). There is also recent evidence that improved methods to measuring formants in speech with high f0 have little advantage when f0 is lower than about 250 Hz (Story and Bunton, 2016), such as in the female speakers in the current study. It is recognized, however, that any formant tracking algorithm applied to naturally produced (and not synthetic) speech is only an approximation. As pointed out by Story and Bunton (2016, p. 94), "(….) the 'true' answer is typically unavailable. That is, an algorithm will deliver measurements of the formant frequencies but whether they are reasonable estimates of the vocal tract resonances produced by the talker is unknown."

To the extent the LPC technique allows conclusions to be drawn, we emphasize that the current results for regional variation support the intermixed automatic-enhanced account of the debated F1/f0 relationship (Honda and Fujimura, 1991; Jacewicz and Fox, 2015a; Van Hoof and Verhoeven, 2011). That is, intrinsic f0 is physiological in nature but regional varieties utilize f0 control in different ways, including regional pitch ranges and variation in f0 shape. Such variations may carry distinct information for listeners/speakers of a given regional variety, cueing their sociocultural identity.

That listeners are particularly sensitive to manipulations in pitch contour in stressed vowels has already been demonstrated in early experiments by Fry (1958). It is highly possible that f0 contours in Southern vowels contribute to the percept of the Southern drawl, supplementing the dominant spectral characteristics of the diphthongal change. While the Southern drawl has always been thought of as belonging to the spectral domain, it is possible that an exaggerated spectral change in stressed vowels is synchronized with a distinct melody of the f0 contour and the mutual influence of the 2 produces the "right" amount of dialect-inherent tune. Conversely, neither the WI nor OH variety studied here carry such a distinct dynamic component, which is also a part of their cultural tune.

We expect the current study to lay a foundation for future investigations of regional variation in the use of f0 in vowels. In particular, the F1/f0 relationship can be further studied both in words produced in isolation and in connected speech to learn more about regional pitch ranges in other varieties of English and in other languages, and about the "intrinsic" aspect of vowel f0 itself. More work is also needed to better understand regional variation in f0 shape in vowels that do not carry the main sentence stress. Also, the speech materials used in the current study did not allow us to inquire into the phonological sources of f0 variation. The complex interactions of segmental effects with higher level prosodic influences still await detailed explorations in the context of regional variation in speech.

At this point, we still need a better understanding of the sources underlying the differential f0 control in regional varieties of American English. The current study suggests that NC speakers use f0 in vowels in a markedly different way, which parallels their distinct use of spectral dynamics and temporal variation in speech. Future studies will need to extend this finding to other speech communities in the South to determine whether a culturally changing environment exerts an influence on f0 variation. This issue also leads to a number of questions related to an active control of pitch patterns in vowels by speakers of different regional varieties and their perceptual awareness of such patterns.

Acknowledgments

Research reported in this paper was supported by the National Institute of Deafness and Other Communication Disorders of the National Institutes of Health under award No. R01DC006871. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank Joseph Salmons for his contributions to this research. Special thanks go to Janaye Houghton and Dilara Tepeli for help with data collection in North Carolina and Wisconsin, respectively. We also thank Jessica Hart, Donna Moore, Alison Davis and Hannah Ortega for their help with data processing.

Disclosure Statement

No conflict of interest existed at the time of submission of the manuscript.

References

▶ Alku P, Pohjalainen J, Vainio M, Laukkanen A-M, Story BH (2013): Formant frequency estimation of high-pitched vowels using weighted linear prediction. J Acoust Soc Am 134:1295–1313.

Arvaniti A, Garding G (2007): Dialectal variation in the rising accents of American English; in Cole J, Hualde J (eds): Laboratory Phonology 9. Berlin, Mouton de Gruyter, pp 547–576.

Atal BS (1975): Linear prediction of speech – recent applications to speech analysis; in Reddy RD (ed): Speech Recognition. New York, Elsevier, pp 221–230.

Beckman ME, Pierrehumbert JB (1986): Intonational structure in Japanese and English. Phonol Yearb 3:255–309.

Boberg C (2005): The Canadian shift in Montreal. Lang Var Change 17:133–154.

▶ Clopper CG, Pisoni D, de Jong K (2005): Acoustic characteristics of the vowel systems of six regional varieties of American English. J Acoust Soc Am 118:1661–1676.

▶ Clopper CG, Smiljanic R (2011): Effects of gender and regional dialect on prosodic patterns in American English. J Phon 39:237–245.

▶ Connell B (2002): Tone languages and the universality of intrinsic F0: evidence from Africa. J Phon 30:101–129.

▶ De Boer B (2011): First formant difference for /i/ and /u/: a cross-linguistic study and an explanation. J Phon 39:110–114.

▶ Derdemezis E, Vorperian HK, Kent RD, Fourakis M, Reinicke EL, Bolt DM (2016): Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups. Am J Speech Lang Pathol 25:335–354.

▶ Deutsch D, Henthorn T, Dolson M (2004): Speech patterns heard early in life influence later perception of the triton paradox. Music Percept 21:357–372.

▶ Deutsch D, Jinghong L, Sheng J, Henthorn T (2009): The pitch levels of female speech in two Chinese villages. J Acoust Soc Am 125:208–213.

▶ Diehl RL, Kluender KR (1989): On the objects of speech perception. Ecol Psychol 1:121–144.

Durian D, Dodsworth R, Schumacher J (2010): Convergence in blue-collar Columbus, Ohio, African American and White vowel systems? In Yaeger-Dror M, Thomas E (eds): AAE Speakers and Their Participation in Local Sound Changes: A Comparative Study. Publications of the American Dialect Society No 94. Durham, Duke University Press, pp 161–190.

▶ Fischer-Jørgensen E (1990): Intrinsic F_0 in tense and lax vowels with specific reference to German. Phonetica 47:99–140.

Fox RA, Jacewicz E (2008): Analysis of total vowel space areas in three regional dialects of American English. Proc Acoustics 08, Paris, pp 495–500.

▶ Fox RA, Jacewicz E (2009): Cross-dialectal variation in formant dynamics of American English vowels. J Acoust Soc Am 126:2603–2618.

▶ Fox RA, Jacewicz E (2017): Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English. J Acoust Soc Am 142:444–459.

▶ Fry D (1958): Experiments in the perception of stress. Lang Speech 1:126–152.

▶ Fulop SA (2010): Accuracy of formant measurement for synthesized vowels using the reassigned spectrogram and comparison with linear prediction. J Acoust Soc Am 127:2114–2117.

Fulop SA (2011): Speech Spectrum Analysis. Berlin, Springer, pp 127–201.

Gordon M (2002): Investigating chain shifts and mergers; in Chambers JK, Trudgill P, Schilling-Estes N (eds): The Handbook of Language Variation and Change. Oxford, Blackwell, pp 244–266.

Halle M, Stevens KN (1971): A note on laryngeal features; in Research Laboratory of Electronics (ed): Quarterly Progress Reports 101. Cambridge, MIT Press, pp 198–213.

Hanson HM (2009): Effects of obstruent consonants on fundamental frequency at vowel onset in English. J Acoust Soc Am 125:425–441.

Hillenbrand JM, Clark MJ, Houde RA (2000): Some effects of duration on vowel recognition. J Acoust Soc Am 108:3013–3022.

Hillenbrand JM, Clark MJ, Nearey TM (2001): Effects of consonantal environment on vowel formant patterns. J Acoust Soc Am 109:748–763.

Hillenbrand JM, Getty LA, Clark MJ, Wheeler K (1995): Acoustic characteristics of American English vowels. J Acoust Soc Am 97:3099–3111.

Hombert JM, Ohala JJ, Ewan WG (1979): Phonetic explanations for the development of tones. Language 55:37–58.

Honda K (1983): Relationship between pitch control and vowel articulation; in Bless D, Abbs J (eds): Vocal Fold Physiology: Contemporary Research and Clinical Issues. San Diego, College-Hill Press, pp 127–137.

Honda K, Fujimura O (1991): Intrinsic vowel F0 and phrase-final F0 lowering: phonological vs biological explanations; in Gauffin J, Hammarberg B (eds): Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms. San Diego, Singular Publishing Group, pp 149–157.

Honda K, Hirai H, Masaki S, Shamada Y (1999): Role of vertical larynx movement and cervical lordosis in F0 control. Lang Speech 42:401–411.

House A, Fairbanks G (1953): The influence of consonant environment upon the secondary acoustical characteristics of vowels. J Acoust Soc Am 22:105–113.

House J, Wichmann A (1996): Investigating peak timing in naturally occurring speech: from segmental constraints to discourse structure; in Hazan V, Rosen S, Holland M (eds): Speech Hearing and Language: Work in Progress 9. London, University College London, pp 99–117.

Jacewicz E, Fox RA (2008a): Amplitude variations in coarticulated vowels. J Acoust Soc Am 123:2750–2768.

Jacewicz E, Fox RA (2008b): The temporal location of rms peak in coarticulated vowels. Proc Acoustics 08, Paris, SFA, pp 627–632.

Jacewicz E, Fox RA (2012): The effects of cross-generational and cross-dialectal variation on vowel identification and classification. J Acoust Soc Am 131:1413–1433.

Jacewicz E, Fox RA (2015a): Intrinsic fundamental frequency of vowels is moderated by regional dialect. J Acoust Soc Am 138:EL405–EL410.

Jacewicz E, Fox RA (2015b): The effects of dialect variation on speech intelligibility in a multitalker background. Appl Psycholinguist 36:729–746.

Jacewicz E, Fox RA (2015c): Eliciting sociophonetic variation in vowel duration (paper ICPHS0016). Proc 18th Int Congr Phon Sci, Glasgow, pp 1–5.

Jacewicz E, Fox RA, Lyle S (2009): Variation in stop consonant voicing in two regional varieties of American English. J Int Phon Assoc 39:313–334.

Jacewicz E, Fox RA, Salmons J (2007): Vowel duration in three American English dialects. Am Speech 82:367–385.

Jacewicz E, Fox RA, Salmons J (2011a): Cross-generational vowel change in American English. Lang Var Change 23:1–42.

Jacewicz E, Fox RA, Salmons J (2011b): Vowel change across three age groups of speakers in three regional varieties of American English. J Phon 39:683–693.

Kent RD, Read C (1992): The Acoustic Analysis of Speech. San Diego, Singular Publishing Group.

Kingston J (1992): The phonetics and phonology of perceptually motivated articulatory covariation. Lang Speech 35:99–113.

Kingston J (2007): Segmental influences on F0: automatic or controlled? In Gussenhoven C, Riad T (eds): Tones and Tunes. Berlin, Mouton de Gruyter, vol 2, pp 171–210.

Kirby JP, Ladd DR (2016): Effects of obstruent voicing on vowel F0: evidence from “true voicing” languages. J Acoust Soc Am 140:2400–2411.

Koops C (2010): /u/-fronting is not monolithic: two types of fronted /u/ in Houston Anglos. University of Pennsylvania Working Papers in Linguistics 16:113–122.

Kurath H (1939): Handbook of the Linguistic Geography of New England. With the Collaboration of Hansen ML, Bloch B, Bloch J. Providence, Brown University.

Labov W (1994): Principles of Linguistic Change. Oxford, Blackwell, vol 1: Internal Factors.

Labov W (2010): Principles of Linguistic Change. Oxford, Blackwell, vol 3: Cognitive and Cultural Factors.

Labov W, Ash S, Boberg C (2006): Atlas of North American English: Phonetics, Phonology, and Sound Change. Berlin, Mouton de Gruyter.

Labov W, Yaeger M, Steiner R (1972): A Quantitative Study of Sound Change in Progress. Philadelphia, US Regional Survey.

Ladd DR (1996): Intonational Phonology. Cambridge, Cambridge University Press.

Ladd DR, Silverman KEA (1984): Vowel intrinsic pitch in connected speech. Phonetica 41:31–40.

Ladd DR, Schepman A, White L, Quarmby LM, Stackhouse R (2009): Structural and dialectal effects on pitch peak alignment in two varieties of British English. J Phon 37:145–161.

Lea WA (1973): Segmental and suprasegmental influences on fundamental frequency contours; in Hyman LM (ed): Consonant Types and Tones. Southern California Occasional Papers in Linguistics No 1. Los Angeles, University of Southern California Press, pp 15–70.

Lehiste I, Peterson GE (1960): Duration of syllable nuclei in English. J Acoust Soc Am 32:693–703.

Lehiste I, Peterson GE (1961): Some basic considerations in the analysis of intonation. J Acoust Soc Am 33:419–425.

Leung K, Jongman A, Wang Y, Sereno J (2016): Acoustic characteristics of clearly spoken English tense and lax vowels. J Acoust Soc Am 140:45–58.

Lindblom B, Agwuele A, Sussman HM, Cortes EE (2007): The effect of emphatic stress on consonant vowel coarticulation. J Acoust Soc Am 12:3802–3813.

Liu L, Shimamura T (2015): Pitch-synchronous linear prediction analysis of high-pitched speech using weighted short-time energy function. J Sig Proc 19:55–66.

Löfqvist A, Baer T, McGarr NS, Story RS (1989): The cricothyroid muscle in voicing control. J Acoust Soc Am 85:1314–1321.

MathWorks Inc, Natick, MA (2014a): MATLAB. http://www.mathworks.com.

Milenkovic P (2003): TF32 software program. Madison, University of Wisconsin.

Mohr B (1968): Intrinsic fundamental frequency variation II. Monthly Internal Memorandum, June. Berkeley, Phonology Laboratory, University of California, pp 23–32.

Monsen RB, Engebretson AM (1983): The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction. J Speech Hear Res 26:89–97.

Morrison GS, Assmann PF (eds) (2013): Vowel Inherent Spectral Change. Berlin, Springer.

Nishio M, Niimi S (2008): Changes in speaking fundamental frequency characteristics with aging. Folia Phoniatr Logop 60:120–127.

Nolan F (2003): Intonational equivalence: an experimental evaluation of pitch scales; Proc 15th Int Congr Phon Sci, Barcelona, pp 771–774.

Ohala JJ, Eukel BW (1987): Explaining the intrinsic pitch of vowels; in Channon R, Shockey L (eds): In Honor of Ilse Lehiste. Dordrecht, Foris, pp 207–215.

Ohde RN (1984): Fundamental frequency as an acoustic correlate of stop consonant voicing. J Acoust Soc Am 75:224–230.

Olive JP, Greenwood A, Coleman J (1993): Acoustics of American English Speech: A Dynamic Approach. New York, Springer.

Peterson GE, Barney HL (1952): Control methods used in a study of the vowels. J Acoust Soc Am 24:175–184.

Pierce CA, Block RA, Aguinis H (2004): Cautionary note on reporting eta-squared values from multifactor ANOVA designs. Educ Psychol Meas 64:916–924.

Pierrehumbert JB, Steele SA (1989): Categories of tonal alignment in English. Phonetica 46:181–196.

Redi LC (2003): Categorical effects in production of pitch contours in English. Proc 15th Int Congr Phon Sci, Barcelona, pp 2921–2924.

Reubold U, Harrington J, Kleber F (2010): Vocal aging effects on F_0 and the first formant: a longitudinal analysis in adult speakers. Speech Commun 52:638–651.

Risdal M, Kohn M (2014): Ethnolectal and generational differences in vowel trajectories: evidence from African American English and the southern vowel system. University of Pennsylvania Working Papers in Linguistics 20:138–148.

Sapir S (1989): The intrinsic pitch of vowels: theoretical, physiological, and clinical considerations. J Voice 3:44–51.

Scicon R and D Inc (2005): PitchWorks. http://www.sciconrd.com.

Shadle CH (1985): Intrinsic fundamental frequency of vowels in sentence context. J Acoust Soc Am 78:1562–1567.

Shadle CH, Nam H, Whalen DH (2016): Comparing measurement errors for formants in synthetic and natural vowels. J Acoust Soc Am 139:713–727.

Silverman KEA, Pierrehumbert JB (1990): The timing of prenuclear high accents in English; in Kingston J, Beckman ME (eds): Papers in Laboratory Phonology 1. Cambridge, Cambridge University Press, pp 495–518.

Sledd J (1966): Breaking, umlaut, and the southern drawl. Language 42:18–41.

Slis IH (1966): A Model for the Distinction between Voiceless and Voiced Consonants. Annual Progress Report 1. Eindhoven, Institute for Perception Research, Eindhoven University of Technology, pp 40–44.

Smiljanic R (2006): Early vs late focus: pitch-peak alignment in two dialects of Serbian and Croatian; in Goldstein L, Whalen DH, Best C (eds): Papers in Laboratory Phonology 8. Berlin, Mouton de Gruyter, pp 495–518.

Stanford JN, Leddy-Cecere TA, Baclawski KP (2012): Farewell to the founders: major dialect changes along the east-west New England border. Am Speech 87:126–169.

Steele SA (1986): Interaction of vowel F_0 and prosody. Phonetica 43:92–105.

Story BH, Bunton K (2016): Formant measurement in children's speech based on spectral filtering. Speech Commun 76:93–111.

Thomas ER (2011): Sociophonetics: An Introduction. Basingstoke, Palgrave Macmillan, pp 184–223.

Titze IR (2000): Principles of Voice Production, ed 2. Iowa City, National Center for Voice and Speech.

Umeda N (1981): Influences of segmental factors on fundamental frequency in fluent speech. J Acoust Soc Am 70:350–355.

Vallabha GK, Tuller B (2002): Systematic errors in the formant analysis of steady-state vowels. Speech Commun 38:141–160.

Van Hoof S, Verhoeven J (2011): Intrinsic vowel F0, the size of vowel inventories and second language acquisition. J Phon 39:168–177.

Watson CI, Harrington J (1999): Acoustic evidence for dynamic formant trajectories in Australian English vowels. J Acoust Soc Am 106:458–468.

Whalen DH, Gick B, Kumada M, Honda K (1998): Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. J Phon 27:125–142.

Whalen DH, Levitt AG (1995): The universality of intrinsic F0 of vowels. J Phon 23:349–366.

Xu Y (1997): Contextual tonal variations in Mandarin. J Phon 25:61–83.

Xu Y (2015): Speech prosody – theories, models and analysis; in Meireles AR (ed): Courses on Speech Prosody. Cambridge, Cambridge Scholars Publishing, pp 146–177.

Xue SA, Hao GJ, Xu L, Moranski T (2008): Speaking fundamental frequency changes in women over time. Asia Pacific J Speech Lang Hear 11:189–194.