Ewa Jacewicz and Robert Allen Fox: JASA Express Letters [http://dx.doi.org/10.1121/1.4934178] Published Online 19 October 2015

# Intrinsic fundamental frequency of vowels is moderated by regional dialect

**Ewa Jacewicz[a) ] and Robert Allen Fox**
*Department of Speech and Hearing Science, The Ohio State University, 1070 Carmack Road, Columbus, Ohio 43210, USA*
*jacewicz.1@osu.edu, fox.2@osu.edu*

**Abstract:** There has been a long-standing debate whether the intrinsic fundamental frequency (IF0) of vowels is an automatic consequence of articulation or whether it is independently controlled by speakers to perceptually enhance vowel contrasts along the height dimension. This paper provides evidence from regional variation in American English that IF0 difference between high and low vowels is, in part, controlled and varies across dialects. The sources of this *F*0 control are socio-cultural and cannot be attributed to differences in the vowel inventory size. The socially motivated enhancement was found only in prosodically prominent contexts.



[AL]
**Date Received:** July 27, 2015 **Date Accepted:** October 5, 2015

## 1. Introduction

There is an apparent (and, possibly universal) relationship between *F*0 values and vowel height: high vowels tend to have a higher *F*0 than low vowels. A long-standing debate has centered on whether this intrinsic fundamental frequency (IF0) pattern is an automatic consequence of vowel articulation or whether it is controlled to exaggerate the differences and enhance vowel contrast. The present study provides new evidence from regional dialect variation suggesting that IF0 is at least partially controlled by the speaker.

The support for IF0 as an automatic aspect of vowel articulation comes primarily from the universality of the effect. An exhaustive review of available published literature in Whalen and Levitt (1995) found these IF0 differences in as many as 31 languages. Using a simple measure of IF0 difference by collapsing across /i, u/ and subtracting the value for /a/, Whalen and Levitt (1995) reported a mean 15.3 Hz-difference across all the languages. Unlike in intonation languages, the IF0 difference appears to be constrained in tone languages, being reduced to single-digit Hz values or even disappearing entirely with low tones (Connell, 2002). There is no consensus in the literature as to the exact sources of IF0 differences. The active mechanism that has been proposed includes the cricothyroid and vocalis muscles although other muscles are also believed to contribute to *F*0 control (Honda *et al.*, 1999). The pull of the tongue on the hyoid bone (the "tongue pull" hypothesis) may also influence IF0 differences (Ohala and Eukel, 1987).

While the mechanism that determines IF0 is still the subject of debate, the issue of whether IF0 is an automatic biomechanical consequence of vowel articulation or whether it represents deliberate effort on the part of the speaker to enhance the vowel height dimension has received a great deal of attention (e.g., Kingston and Diehl, 1994; Whalen and Levitt, 1995; Whalen *et al.*, 1998; Kingston, 2007). Accordingly, if IF0 is a purely phonetic effect, then it should be found in all contexts of vowel articulation, being "resistant" to prosodic and consonantal influences. But if IF0 is a phonological enhancement feature which is implemented to perceptually enhance vowel contrast, then it may be actively controlled by the speakers in order to "exaggerate" the contrast. Consequently, the magnitude of IF0 difference between high and low vowels is expected to be greater in prosodically prominent contexts such as in accented syllables and it may be reduced or even disappear completely in unaccented syllables (Ladd and Silverman, 1984; Kingston, 2007).

More recently, Van Hoof and Verhoeven (2011) proposed a mixed physiological-enhancement account, which assumes that IF0 is a physiologically determined aspect of vowel articulation but one that can be, in part, controlled by the

[a)]Author to whom correspondence should be addressed.

speaker. Of relevance to the current study is the evidence provided in Van Hoof and Verhoeven (2011) from second language (*L*2)-accented productions. In particular, IF0 was measured in native speakers of a language with a small vowel system (Arabic, 3 vowels), with a large system (Dutch, 12 vowels) and in *L*2 Dutch spoken by native-Arabic speaking learners. There was a relation between a vowel inventory size and IF0: while present in all three varieties, the difference in IF0 between the high vowels [i:, u:] and the low vowel [a:] was significantly larger in Dutch than Arabic when produced by native speakers but did not differ significantly between Arabic and Dutch when produced by *L*2 speakers. This result was interpreted as indicating that languages with large vowel systems may exploit IF0 differences to enhance vowel contrasts, and this intrinsic language-specific difference in *F*0 for high vs low vowels must be learned by *L*2 speakers whose native language has a small vowel inventory. If so, then IF0 is not simply a consequence of vowel articulation but a language-specific feature which is controlled and, at least in part, acquired separately.

This possibility leads to a question related to IF0 in the context of regional dialect variation. In particular, if vowels in regional dialects of the same language differ in their spectral and temporal characteristics (Clopper *et al.*, 2005; Jacewicz *et al.*, 2010, 2011), do they also differ in their IF0? Stated differently, if IF0 is in part controlled by a separate mechanism, is this mechanism also available for fine-grained manipulation across regional varieties of the same language?

## 2. Methods

### *2.1 Participants*

Thirty-six women participated. All were long-time residents of 1 of 3 dialect regions in the United States: 12 were from Western North Carolina (NC) (Jackson County) and spoke the local Southern variant of American English, 12 were from Central Ohio (OH) (Columbus and suburbs) and spoke the Midland variety, and 12 were from Southeastern Wisconsin (WI) (Madison area) and spoke the Midwestern English. These participants were selected from a larger pool of 48 speakers used in Fox and Jacewicz (2009) for vowel formant analysis. The current reduction in the number of participants was due to extensive creaky phonation in some speakers which could compromise the accuracy of *F*0 measurements. The mean ages in years (and standard deviations) for each dialect group were 56.5 (3.2) for NC, 57.6 (5.9) for OH, and 58.2 (4.0) for WI. A one-way analysis of variance (ANOVA) indicated that these slight age differences across groups were not significant. The choice of the middle-aged participants as producing the most representative dialect-specific sample was guided by recent evidence about cross-generational sound change in each of these three speech communities (Jacewicz *et al.,* 2011).

### *2.2 Stimuli and procedures*

To measure the effects of prosodic and consonantal contexts on the difference in IF0 between the high and low vowels, the target stimuli controlled the degree of emphasis (stressed, unstressed) and post-vocalic obstruent voicing (voiced, voiceless). According to several reports, obstruent voicing can influence the *F*0 of the preceding vowels with voiceless consonants raising their *F*0 and voiced consonants lowering their *F*0 (e.g., Slis, 1966). The speech materials utilized in this study came from a larger cross-dialectal corpus and the experimental procedures were detailed in Fox and Jacewicz (2009). Briefly, the five vowels /ɪ, ɛ, e, æ, aɪ/ were studied in that corpus which 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 word that followed was always "are." There were two types of declarative sentences in which (1) the target word carried the main sentence stress and the prominent vowel was accented (e.g., Ted thinks the fall **BIDS** are low) and (2) the target word occurred in unstressed position and the vowel was unaccented (e.g., Ted thinks the fall **bids** are LOW). A contrastive stress paradigm was used to elicit the two types of vowel production. The participant read each sentence fluently and without pauses, placing the main sentence stress on the word in all caps and the target words (indicated in bold here for the sake of exposition but not in the original text) were used in the acoustic analysis. The sentences were presented on a computer monitor in random order. A head-mounted Shure SM10A dynamic microphone (Shure Inc., Niles, IL) was positioned about 1.5 in. from the speaker's lips. The read sentences were recorded and digitized at a 44.1-kHz sampling rate with 16-bit quantization directly onto a hard disk drive.

For the current study, only two of the five vowel categories were analyzed, /ɪ/ and /æ/, representing a high and a low vowel, respectively. Because no back vowels were included in the stimulus set, the current study only focuses on the front vowel dimension. For these two vowels, 24 sentences were obtained from each speaker except for 1 who produced half of the set. Out of the 24 sentences, 16 were produced with the target word stressed (2 vowels × 2 consonant contexts × 4 repetitions) and 8 with the target word unstressed (2 repetitions instead of 4). From this data set, 33 target words (3.87%) were excluded due to creaky voice throughout the vowel, mostly in unstressed positions. Altogether, a total of 819 tokens from 36 speakers were analyzed in the study, 413 for /ɪ/ and 406 for /æ/, of which 554 were stressed and 265 were unstressed.

### 2.3 Acoustic analysis

Prior to acoustic analysis, the digitized tokens were down sampled to 11.025 kHz and low-pass filtered at 1 kHz. *F*0 analysis was based on *F*0 tracks throughout the vowel (representing *F*0 contour shapes). *F*0 tracks were computed using autocorrelation implemented in MATLAB in a series of 16-ms wide measurement windows (with 50% overlap) over the course of the vowel. Vowel onsets and offsets were located by hand and defined using standard segmentation criteria (details are provided in Fox and Jacewicz, 2009). Several custom MATLAB routines were written to facilitate *F*0 analyses and ensure accuracy of *F*0 measurement. Mistracked *F*0 values were examined and hand-corrected after consulting outputs of the commercially available programs TF32 (Milenkovic, 2003) and PitchWorks (Scicon R&D, 2005) and recalculating the questionable *F*0 tracks using a separate MATLAB program. A reliability check was done on all tokens by R. A. Fox and corrections, if necessary, were made prior to data processing in spreadsheets. *F*0 measures included overall mean *F*0, onset *F*0, offset *F*0, and peak *F*0 in stressed vowels. The initial *F*0 value, onset *F*0, was that obtained at the first measurement window following vowel onset. Peak *F*0 (or *F*0 maximum) was the highest detected *F*0 value in a stressed vowel. The offset *F*0 was the *F*0 value obtained for the last measurement window.

## 3. Results

The IF0 differences between /ɪ-æ/ were first examined using a repeated-measures ANOVA with the within-subject factors *F*0 measure (onset, peak, offset, and overall *F*0), stress and coda voicing, and the between-subject factor dialect. The two main effects of *F*0 measure and coda voicing were not significant, indicating that the IF0 difference between /ɪ-æ/ did not differ as a function of either measurement location or whether the postvocalic consonants were voiced or voiceless. The main effect of stress was significant [$F(1, 33) = 92.42$, $p < 0.001$, $\eta_p^2 = 0.737$]; the IF0 difference between /ɪ-æ/ was close to zero for unstressed vowels ($M = 0.6$ Hz) but far greater for the stressed vowels ($M = 27.6$ Hz). The main effect of dialect was also significant [$F(2, 33) = 5.39$, $p = 0.001$, $\eta_p^2 = 0.366$]. Scheffe's test showed that NC IF0 difference ($M = 5.6$ Hz) was significantly smaller than either OH ($M = 16.3$ Hz) or WI ($M = 20.5$ Hz) while the latter two did not differ from one another. Significant interactions arose between *F*0 measure × stress and *F*0 measure × voicing but they did not yield consistent patterns and are not discussed. No other interactions were significant.

Figure 1 displays the IF0 values for /ɪ/ and /æ/ for each *F*0 measure in stressed vowels before a voiced coda. The same basic overall pattern also occurred before a voiceless coda and is not shown here. As is evident, *F*0 for the high vowel /ɪ/ is consistently higher than for the low vowel /æ/ in all three dialects. This finding supports the universal aspect of IF0 as a feature which is physiologically determined. However, the magnitude of the difference (ΔF0) is dialect-specific: it is smallest in NC and comparatively greater in OH and WI. These variations suggest that dialects differ in their use of IF0 in vowels.

Shown in Fig. 2 is a corresponding set of *F*0 values for unstressed vowels followed by a voiced coda. IF0 differences across all *F*0 measures are negligible and dialectal differences have been mostly eliminated. It is of note that in the Southern NC dialect *F*0 in /ɪ/ is somewhat lower than *F*0 in /æ/ although this is a small difference. Overall, these results support the position that IF0 differences between high and low vowels are exaggerated in prominent prosodic contexts and diminish in unstressed vowels. Importantly, the size of the enhancement is dialect-specific and regional dialects appear to control the use of *F*0 in those prominent positions.

The next set of analyses explored the relationship between *F*0 and vowel height to establish the correspondence between dialect-specific IF0 difference and dialect-specific difference along the height dimension. First, an ANOVA for *F*1 difference

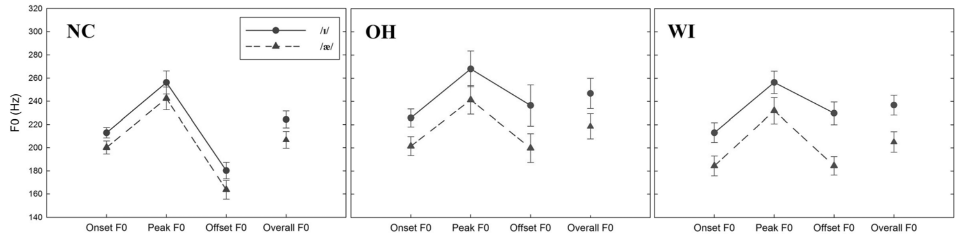


Fig. 1. Mean IF0 difference between high /ɪ/ and low /æ/ stressed vowels before a voiced coda measured at onset, peak and offset *F*0 and as overall *F*0 for OH, WI, and NC speakers. The error bars represent one standard error.

between /ɪ-æ/ was used with the within-subject factors stress and voicing and the between-subject factor dialect. The values for *F*1—indicative of vowel height—were synchronized with *F*0 measurements at peak *F*0, primarily to be compatible with previous studies which typically sampled *F*0 and *F*1 "in the vowel's middle third portion" (Van Hoof and Verhoeven, 2011) or "steady-state" (Hillenbrand *et al.*, 1995). Details about the formant analysis method can be found in Fox and Jacewicz (2009). The main effect of stress was not significant because the stress-related changes were primarily in *F*2 and not in *F*1. The main effect of voicing was significant [$F(1, 33) = 26.33$, $p < 0.001$, $\eta_p^2 = 0.444$]; the difference in *F*1 between /ɪ-æ/ was greater before a voiceless coda ($M = 317.5$ Hz) than before a voiced coda ($M = 275.3$ Hz). While the current tendency for *F*1 values to be lower in the context of voiced consonants is consistent with the literature (Hillenbrand *et al.*, 2001), the smaller difference between /ɪ-æ/ before a voiced coda relative to a voiceless coda was due to a larger downward *F*1 shift for /æ/ than for /ɪ/. This pattern was consistent for all three dialects. The main effect of dialect was significant [$F(2, 33) = 3.30$, $p = 0.049$, $\eta_p^2 = 0.167$] and Scheffe's test showed that WI difference in *F*1 was significantly smaller ($M = 250.5$ Hz) than either NC ($M = 311.5$ Hz) or OH ($M = 327.2$ Hz), which did not differ from one another. There were no significant interactions.

Next, the relationship between the /ɪ-æ/-difference in IF0 and between the /ɪ-æ/-difference in *F*1 was explored using correlation analysis. Pearson's correlations were run for stressed voiced, stressed voiceless, unstressed voiced, and unstressed voiceless combinations. None of the correlations was significant (the highest Pearson's *r* was 0.322 for the unstressed voiced combination). Clearly, there was no meaningful relationship between size of IF0 difference and the size of *F*1 difference between /ɪ/ and /æ/. In particular, the stress-related variations in IF0 did not correspond to the variations in *F*1 and the significant changes in *F*1 due to consonant voicing did not match up with a lack of voicing-related changes in IF0. Finally, the dialect-specific variation in IF0 was quite different from that in *F*1. In particular, the /ɪ-æ/-difference in IF0 was smallest in NC and greatest in WI whereas the /ɪ-æ/-difference in *F*1 showed the reverse pattern.

## 4. Discussion

This study aimed to determine whether, in addition to cueing vowel height, IF0 represents a dialect-specific feature which can be controlled to some extent by speakers. This question was motivated by the proposal that the size of the IF0 difference is language-specific and is influenced by vowel inventory size (Van Hoof and Verhoeven, 2011). In the Van Hoof and Verhoeven study, the smaller IF0 difference between high and low vowels in Arabic (a language with 3 vowels) was "transferred" to *L*2 Dutch, but the IF0 difference in Dutch spoken by native speakers (a language with 12 vowels) was significantly larger. This result was interpreted as indicating that the size of the IF0 difference is related to the size of the vowel inventory and that speakers

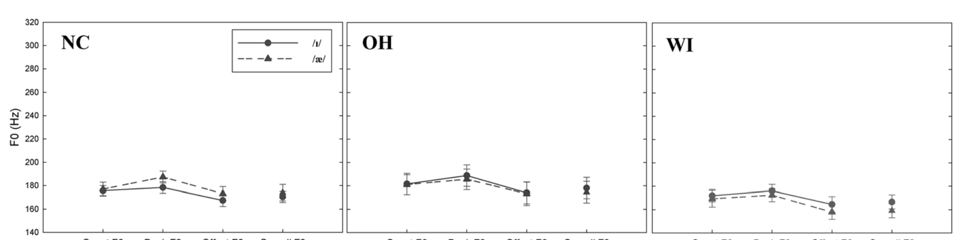


Fig. 2. Mean IF0 difference between high /ɪ/ and low /æ/ unstressed vowels before a voiced coda. The error bars represent one standard error.

deliberately use IF0 to maximize (or perceptually enhance) vowel contrast. The current study provides evidence that the IF0 difference between high and low vowels can also vary as a function of dialect of the same language, which cannot be attributed to differences in the vowel inventory size. Rather, the sources of the control of *F*0 use must be socio-cultural, reflecting regional variation in speech.

The current study also found that these deliberate efforts were manifested only in prominent prosodic contexts such as to convey semantic emphasis (focus) in a sentence. Any *F*0 differences between high and low vowels were diminished although they did not entirely disappear when the vowels were unstressed, consistent with previous reports (e.g., Ladd and Silverman, 1984). In addition, no meaningful relationship between the size of the IF0 difference and the size of the *F*1 difference between high and low vowels was found. This evidence leads to the conclusion that the quality of vowels and *F*0 control are two aspects that need to be learned separately.

The small IF0 difference in NC dialect ($M = 5.6$ Hz) is particularly intriguing because it is certainly within the range of values found for tone languages. The values in either OH ($M = 16.3$ Hz) or WI ($M = 20.5$ Hz) are more consistent with those for intonation languages, including American English. We do not speculate on this curious outcome at present due to the preliminary nature of the study. Undoubtedly, the current investigation is limited in scope as only front vowels were examined and the data come from female speakers only. There is some indication that there can be a difference in the size of IF0 by sex (Whalen and Levitt, 1995) although the effect may be marginal (Hillenbrand *et al.*, 1995). We wish to underscore the validity of the current cross-dialectal IF0 differences, however. It is the case that elicitation methods play an important role in characterizing *F*0 variation. Consequently, IF0 difference is expected to be smaller in isolated words than for vowels bearing the main sentence stress. The validity of the current results stems from measurements obtained following a common experimental protocol, which has reduced the large variability found across other studies due, in large part, to differences in design (see Whalen and Levitt, 1995).

The current findings underscore the complexity of *F*0 control, which can be influenced not only by sentence prosody but also by a range of socially-determined factors. There is a possibility that regional dialects employ typical dialect-specific *F*0 values and *F*0 ranges which reflect a differential use of speaking *F*0 across speech communities. If speech communities indeed differ in their prevailing pitch ranges that they use for speech as has been proposed elsewhere (Deutch *et al.*, 2009), then these cross-dialectal differences could also appear in IF0 values. Presumably, the dialect-specific IF0 differences such as observed in this study would be deliberately introduced into the articulation system to enhance the perception of social identity, predominantly in contexts of high prosodic prominence. This possibility needs to be verified in future work.

**Acknowledgments**

Research reported in this paper was supported by NIDCD Grant No. R01DC006871. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH. We thank Joseph Salmons for his contributions to this research.

**References and links**

Clopper, C. G., Pisoni, D., and de Jong, K. (**2005**). “Acoustic characteristics of the vowel systems of six regional varieties of American English,” J. Acoust. Soc. Am. **118**, 1661–1676.

Connell, B. (**2002**). “Tone languages and the universality of intrinsic F0: Evidence from Africa,” J. Phonetics **30**, 101–129.

Deutsch, D., Jinghong, L., Sheng, J., and Henthorn, T. (**2009**). “The pitch levels of female speech in two Chinese villages,” J. Acoust. Soc. Am. **125**, EL208–EL213.

Fox, R. A., and Jacewicz, E. (**2009**). “Cross-dialectal variation in formant dynamics of American English vowels,” J. Acoust. Soc. Am. **126**, 2603–2618.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). “Effects of consonantal environment on vowel formant patterns,” J. Acoust. Soc. Am. **109**, 748–763.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). “Acoustic characteristics of American English vowels,” J. Acoust. Soc. Am. **97**, 3099–3111.

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

Jacewicz, E., Fox, R. A., and Salmons, J. (**2011**). “Cross-generational vowel change in American English,” Lang. Var. Change **23**, 45–86.

Jacewicz, E., Fox, R. A., and Wei, L. (**2010**). “Between-speaker and within-speaker variation in speech tempo of American English,” J. Acoust. Soc. Am. **128**, 839–850.

Kingston, J. (**2007**). “Segmental influences on F0: Automatic or controlled?,” in *Tones and Tunes*, Vol. 2, edited by C. Gussenhoven and T. Riad (Mouton de Gruyter, Berlin, Germany), pp. 171–210.

Kingston, J., and Diehl, R. (**1994**). “Phonetic knowledge,” Language **70**, 419–454.

Ladd, D. R., and Silverman, K. E. A. (**1984**). "Vowel intrinsic pitch in connected speech," Phonetica **41**, 31–40.
Milenkovic, P. (**2003**). TF32 software program (University of Wisconsin, Madison, WI).
Ohala, J. L., and Eukel, B. W. (**1987**). "Explaining the intrinsic pitch of vowels," in *In Honor of Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris Publications, Dordrecht), pp. 207–215.
Scicon R&D, Inc. (**2005**). PitchWorks. http://www.sciconrd.com (Last viewed September 24, 2015).
Slis, I. H. (**1966**). "A model for the distinction between voiceless and voiced consonants," *Annual Progress Report* 1, Institute for Perception Research, Eindhoven University of Technology, Eindhoven, The Netherlands, pp. 40–44.
Van Hoof, S., and Verhoeven, J. (**2011**). "Intrinsic vowel F0, the size of vowel inventories and second language acquisition," J. Phonetics **39**, 168–177.
Whalen, D. H., Gick, B., Kumada, M., and Honda, K. (**1998**). "Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0," J. Phonetics **27**, 125–142.
Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic F0 of vowels," J. Phonetics **23**, 349–366.