

SOCIOCULTURAL INFLUENCES ON FUNDAMENTAL FREQUENCY IN VOWELS: THE CASE OF AFRICAN AMERICAN ENGLISH

Yolanda Feimster Holt^a, Ewa Jacewicz^b, and Robert A. Fox^b

^aDept. of Communication Sciences & Disorders, East Carolina University, Greenville, NC, USA

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

holty@ecu.edu, jacewicz.1@osu.edu, fox.2@osu.edu

ABSTRACT

Indisputably, high vowels have higher intrinsic fundamental frequency (If0) than low vowels. An unresolved question is whether this If0 is a purely automatic by-product of vowel articulation, or whether it can be actively controlled by speakers to enhance vowel contrast. A third mixed position proposes that the effect is physiological but the magnitude of If0 difference between high and low vowels is language-specific and can be controlled. Recently, the size of If0 difference was shown to vary with regional dialect. Further testing this mixed physiological-enhancement hypothesis, the current study found the distinctive f0 control in an ethnolect. Data were collected in two regionally distinct African American English communities. The magnitude of If0 difference was equal in both groups irrespective of different f0 values associated with common F1. This f0/F1 mismatch indicates that If0 in vowels can be controlled, possibly serving as a marker of sociocultural identity and group membership.

Keywords: sociophonetics, vowels, fundamental frequency, African American English, dialect.

1. INTRODUCTION

A well-documented and possibly universal aspect of vowel production is that high vowels typically have higher fundamental frequency (f0) than low vowels [34]. Although the exact mechanism that determines this intrinsic f0 (If0) pattern is still not well understood, a long-standing debate has focused on whether If0 is a purely phonetic effect reflecting an automatic biomechanical consequence of vowel articulation [27,11,33], or whether it can be actively controlled by the speaker to perceptually enhance the contrast between a high and a low vowel [4,17,19]. The primary evidence for the latter position comes from findings that the If0 *difference* between high and low vowels is greater in prosodically prominent contexts such as in stressed syllables and is reduced or even completely neutralized when syllables are unstressed [21,18].

As a third possibility, it has been proposed that If0 may be physiological (and automatic) in nature but it

may also be actively controlled by the speaker to improve or enhance certain language-specific vowel characteristics [10]. This mixed physiological-enhancement account was further supported by the finding that native speakers of one language can “transfer” their native If0 to their second language [32], which implies that the magnitude of the If0 difference between high and low vowels is a language-specific feature that can be, at least in part, under speaker control.

The most recent explorations of the mixed physiological-enhancement proposal revealed that the size of the If0 difference can also vary in regional varieties of the same language [13, 14]. In particular, systematic If0 differences were found in stressed vowels in three distinct varieties of American English (AE); these differences were diminished when the vowels were unstressed [13]. In a follow-up study, regional If0 differences were found not only in stressed vowels in connected speech, but also in vowels in citation-form hVd-syllables produced as isolated items [14]. Importantly, there was no consistency between the magnitude of If0 difference and the magnitude of F1 difference across dialects. That is, in one dialect, the F1 difference between a high and a low vowel was large and the corresponding If0 difference was small, whereas the reverse was true for another dialect. Mismatches of this type suggest that, although the *intrinsic* aspect of f0 variation in vowels may be common to all regional varieties, the magnitude of If0 difference is dialect-specific. Presumably, f0 control can be learned separately, perhaps as a part of a local cultural tune [31], and can communicate sociocultural identity in concert with other regional features such as distinct pitch ranges [3] and distinct tempo of speech [15].

The current study further inquired into the sociocultural aspect of the If0 difference. We conjectured that, assuming that f0 control can be learned separately, the distinctive use of f0 can also be manifested in an ethnolect. This is because markers of sociocultural ethnic (racial) identity tend to be strong and distributed over an array of linguistic structures that, as a whole, cohere for the ethnic group. To examine this possibility, we chose African American English (AAE), an intensely studied ethnicity-based social variety of AE with its own

system of semantic, morphosyntactic, prosodic, and phonological rules [22,35,30]. The current understanding of AAE in the United States is that certain defining features of the variety are relatively stable and uniform across the country (e.g., the reduction of the temporal contrast between tense and lax vowels, systematic variation in word final consonant voicing, or distinctive use of vowel duration [28,7,8]), but the ethnolect also shares some features of AE, both mainstream and regional, such as the use of f_0 declination in read speech and participation in regional vowel shifts [6, 9]. We hypothesized that, if f_0 control contributes to the expression of socioculturally based ethnic identity, then the magnitude of the f_0 difference between high and low vowels will be common across AAE communities in different parts of the US.

To test this hypothesis, the data were collected in two different AAE communities in the southern state of North Carolina (NC). The AAE participants were lifelong residents of either mountainous western NC (WNC) or coastal Eastern NC (ENC), areas separated by 300 miles (482 km). Of relevance, the vowel systems in these two communities have been differentially influenced by regional variation in AE. The AAE in WNC shows evidence of the Southern Shift, a well-known set of vowel changes affecting most of the American South [20]. The AAE in ENC, however, is not affected by this shift, and exhibits typical AAE markers such as a lack of back vowel fronting for /u/ and /o/. Consistent with core features of the broad Southern AE spoken in the southern states, AAE speakers in ENC and WNC pronounce the diphthong /ai/ as a long monophthong [ɑ:].

Our hypothesis will be supported if the AAE speakers in ENC and WNC do not differ in their use of f_0 to convey the contrast between high and low vowels (i.e., the f_0 difference will be common to both). In terms of the f_0 and F1 relationship, there are three possible scenarios. First, f_0 and F1 values of high and low vowels will be the same in ENC and WNC and thus the f_0 difference and the corresponding F1 difference will be common. Second, there can be a community-specific shift in f_0 /F1 so that, being inversely related, lowering of the vowels in the acoustic space (manifested as an increase in F1) will correspond to a decrease in f_0 (and vice versa). Consequently, the magnitudes of the f_0 and F1 differences can still be common despite the distinct values of f_0 and F1 in ENC and WNC. These two scenarios will indicate that f_0 is an automatic effect of vowel articulation. The third possibility is that there will be a mismatch between f_0 and F1. In particular, f_0 difference can still be the same in both communities but the two will differ in their use of F1 to mark a distinction between high and low vowels.

The reverse is also possible in that F1 will be shared but the communities will differ in their use of f_0 . The third scenario (i.e., the mismatch) will indicate that f_0 control can be learned separately.

2. METHODS

2.1. Participants

For the current study, productions of 19 AAE middle-age women ranging in age from 30 to 50 years were analyzed, 8 from WNC and 11 from ENC. Although more speakers participated in the experiment, their productions had to be excluded from f_0 analyses due to the detrimental effects of creaky voice, colds, or smoking on the accuracy of f_0 measurements. The women in WNC resided in Iredell County (mean age = 47.1 years) and those in ENC lived in Pitt County (mean age = 38.0 years). All participants had at least high school education and none had more than a professional graduate (Masters) degree.

2.2. Speech material and analysis

A full set of 14 AE vowels was obtained from each participant. The vowels were produced in hVd-frame as citation-form tokens. The randomized hVd prompts appeared on a computer monitor and the participant read each item, one at a time, providing 3 repetitions of each. The tokens were recorded and digitized at a 44.1-kHz sampling rate with 16-bit quantization.

For the analysis of the f_0 difference between high and low vowels, we followed procedures outlined in [14]. The vowel /u/ in *who'd* (rather than /i/ in *heed*) was selected as the high vowel because f_0 in /u/ was higher than f_0 in /i/, which is also consistent with previous findings [34, 14]. Given that in this southern variety of AAE the monophthongal /ai/ in *hide* was lower in height and had lower f_0 than the vowel /a/ in *hod*, we selected /ai/ (pronounced as [ɑ:]) as the lowest monophthongal vowel in the AAE system. This was also the case for the Southern AE variety studied in [14]. In the remainder of this paper, we will refer to this monophthongal /ai/ as /a/, bearing in mind that this vowel represents the lowest monophthong in Southern AE.

The selected tokens were first downsampled to 11.025 kHz and low-pass filtered at 1 kHz. Vowel onsets and offsets were located by hand and the f_0 tracks over the vowel were computed using autocorrelation method in a series of 16-ms windows with 50% overlap. These computations were implemented in MATLAB [24]. The f_0 values selected for the analyses of f_0 difference were obtained at the temporal point corresponding to 35% of vowel's duration. Measuring f_0 early in the vowel, before the

midpoint, is the recommended approach for vowels produced in isolated syllables [34, 14]. Visual inspection of f0 tracks ensured that the measurements were taken prior to f0 fall. The corresponding F1 values were obtained from LPC spectra at the same temporal location in the vowel. We started with 14 coefficients and made adjustments to both analysis bandwidth and the number of coefficients as needed on a speaker-by-speaker basis. This was done using TF32 software package [26]. The measurements of f0 from the tracks generated in MATLAB were then compared with f0 tracks generated using autocorrelation in TF32 and PitchWorks [29] and hand corrections were made as necessary. Reliability check was done on all tokens.

Given that the study controlled for gender and age, and also to facilitate comparisons with other published reports, f0 and F1 values were not normalized across participants. Statistical linear mixed-effects analyses were carried out in SPSS v. 25 [12]. For each dependent variable (analyses were based on individual productions of each token), a baseline model only included the intercept and community (WNC, ENC) was then entered as a fixed effect. Participant was a random effect. The significance of the fixed effect was based on likelihood ratio tests and associated *p*-values.

3. RESULTS

Figure 1 shows the group variance for If0 difference (left panel) and for the corresponding F1 difference (right panel) between /u/ and /a/ for WNC and ENC speakers. Boxplots show group median, inter-quartile range (box edges) and the 10th and 90th percentiles (whisker edges). Group mean is marked in red (dashed line). Statistical results support the visual impression that the two AAE communities did not differ in either If0 difference or F1 difference. The analysis of the If0 difference failed to show the effect of community ($\chi^2(1) = 0.009, p = .920$). Likewise, the analysis of F1 difference failed to show the effect of community ($\chi^2(1) = 0.479, p = .490$). These results demonstrate no differences between WNC and ENC. We interpret them as indicating that the magnitude of If0 difference and the corresponding magnitude of F1 difference are a common property of AAE spoken in these two communities.

The f0/F1 relationship was further examined in separate analyses of f0 and F1 for /u/ and /a/. Group-averaged f0 and F1 are displayed in Figure 2. We observe that ENC speakers (top, in red) produced both vowels with a higher f0 than WNC speakers (bottom, in blue) whereas F1 values did not seem to differ between the groups. Linear mixed effects analyses carried out on f0 confirmed that community

was a significant predictor of f0 for both /u/ ($\chi^2(1) = 6.731, p = .010$) and /a/ ($\chi^2(1) = 7.078, p = .010$), with ENC speakers producing the vowels with a higher f0 (23 Hz for either /u/ or /a/) than WNC speakers. The analyses of F1 revealed no significant differences in vowel height as a function of community for either /u/ ($\chi^2(1) = 1.937, p = .160$) or /a/ ($\chi^2(1) = 1.198, p = .270$).

Figure 1: If0 and F1 differences between /u/ and /a/

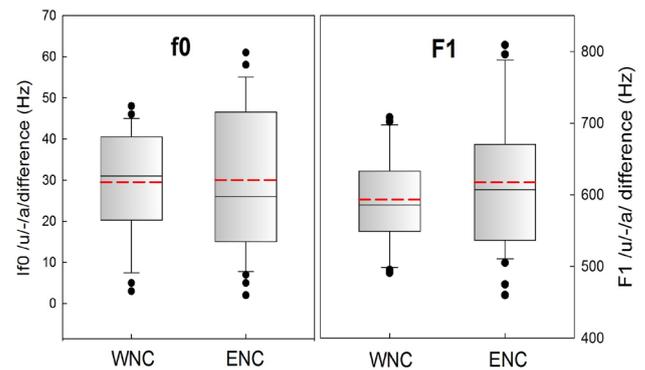
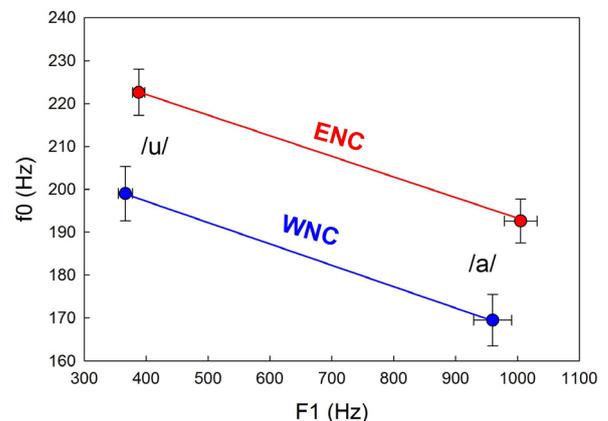
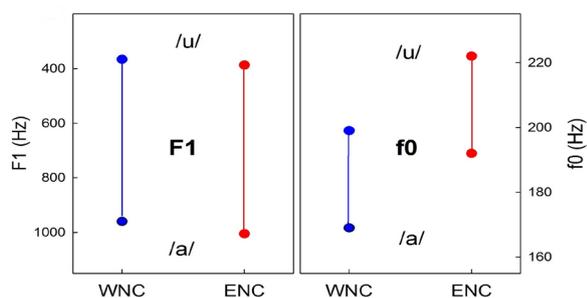


Figure 2: Average (s.e.) f0 and F1 for /u/ and /a/.



Importantly, F1 values in ENC speakers, although not statistically different from WNC, did not, in absolute terms, correspond to their comparatively higher f0. As shown in Figure 3 (redrawn from Figure 2), F1 values for both /u/ and /a/ in ENC speakers were only slightly higher relative to WNC speakers but their f0 values were significantly higher. The differences in F1 between the groups were too small to reach statistical significance (21 Hz for /i/ and 45 Hz for /a/) but the lack of directional agreement between f0 and F1 is striking. Mismatches of this kind suggest at least some active involvement of the speakers in executing the F1/f0 relationship, supporting the proposition that f0 control can be learned separately.

Figure 3: F1/f0 relationship in /u, a/.



4. DISCUSSION

The results of the current study support our conjecture that the magnitude of If0 difference between high and low vowels may be linked to sociocultural variation in speech. We measured f0 and F1 in AAE speakers in two distant and regionally distinct AAE communities and found that the magnitude of If0 difference was equal in both groups, and that it was unaffected by regional influences on their respective vowel systems. This resistance suggests that the size of If0 difference may be a shared feature of the ethnolect, at least for these two varieties spoken in NC, reflecting ethnic (racial) identity of its speakers.

The finding that F1 difference between high and low vowels was also common in both groups could be misinterpreted as indicating that f0/F1 relationship is automatic and predetermined by biomechanics of vowel production. The separate analyses of f0 and F1 showed, however, that vowels can be produced with higher f0 (i.e., at a higher pitch) but this higher f0 may not be associated with lower F1. In particular, the two groups differed in their f0 but their F1 values were not statistically different; there was even a slight increase in F1 for vowels produced with higher f0 in ENC speakers when compared with WNC speakers. This f0/F1 mismatch suggests that f0 use in vowels can be controlled and may reflect a regional “tune,” that is, a specific range of speaking f0 associated with a particular regional group or sociocultural speech community [3].

Support for this position comes from the recent emerging evidence that the magnitude of If0 difference between high and low vowels can be altered by regional variation in AE [13,14]. Measured at a point of high energy in hVd-tokens, closer to vowel onset (at the 35% point), the reported If0 difference in female speakers was about 24 Hz in AE variety spoken in western NC, 32 Hz in central Ohio, and 43 Hz in southern Wisconsin [14, their Figure 2]. These distinct f0 ranges did not correspond to the magnitude of F1 differences in these AE dialects, which were 635 Hz, 549 Hz and 594 Hz, respectively. In the current study, the If0 difference was 30 Hz for each AAE community and the F1 differences were very close, 593 Hz in WNC and 617 Hz in ENC. This

comparison positions NC AAE speakers as having an If0 difference similar to Ohio speakers and an F1 difference similar to Wisconsin speakers, which creates yet another combination of If0 and F1.

The apparent mismatches between If0 and F1 differences reflect no agreement between positional variation in high and low vowels in regional vowel systems (such as related to chain shifts) and f0 use. We can thus admit the possibility that regional and ethnic varieties may choose a particular pitch range, a specific f0 span, as one of the variables cuing their socio-cultural identity and group membership.

This possibility is not implausible if we consider that human voice (and thus, f0) conveys a wealth of information about speaker characteristics, including age, gender, health, education or social status so that different social groups may use f0 differently [5]. The typical f0 range deployed in spoken language by a specific subgroup or population has been referred to as speaking f0 or SFF [1]. Studies exploring speaking f0 across languages reported significant differences, suggesting that different languages may use different f0 ranges and typical language-specific f0 values as a part of their phonetic structure. Cross-language differences in SFF range were found for Polish vs. AE [23], British English vs. German [25], AE vs. Mandarin [16], among many others.

Still little is known whether differences in SFF exist across regional and ethnic varieties of the same language. It has been proposed that speakers of a particular variety 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) influences their speech production; this mental representation is then utilized when acquiring a second language or dialect [2,3]. Consequently, if speakers of different regional and ethnic varieties utilize specific f0 values and f0 ranges, then these cross-varietal differences should also emerge in vowel production. We can thus reasonably expect that production of experimental syllables or words may be influenced by speakers’ internal representation of pitch ranges typical of their speech communities.

In conclusion, the common magnitude of If0 difference between high and low vowels found in AAE speakers living in two distinct and distant communities suggests that this parameter may reflect a shared feature of the ethnolect. The finding that the two communities use different f0 values associated with the same F1 indicates that If0 in vowels can be controlled by speakers, supporting the physiological-enhancement hypothesis advanced in previous work [10,13,14,32]. Future research probing different speech communities and including more complex and tightly controlled speech tasks will ascertain the degree to which the present findings can be generalized.

5. REFERENCES

- [1] Baken, R. J., Orlikoff, F. F. 2000. *Clinical Measurement of Speech and Voice, 2nd ed.* San Diego, CA: Singular Publishing Group, 145-224.
- [2] 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.
- [3] Deutsch, D., Jinghong, L., Sheng, J., Henthorn, T. 2009. The pitch levels of female speech in two Chinese villages. *J. Acoust. Soc. Am.* 125, EL208–EL213.
- [4] Diehl R. L., Kluender K. R. 1989. On the objects of speech perception. *Ecol. Psychol.* 1, 121-144.
- [5] Hollien, H., Hollien, P. A., de Jong, G. 1997. Effects of three parameters on speaking fundamental frequency. *J. Acoust. Soc. Am.* 102, 2984-2992.
- [6] Holt, Y. 2018. Mechanisms of vowel variation in African American English. *J. Speech, Lang., Hear. Res.* 61, 197-209.
- [7] Holt, Y., Jacewicz, E., Fox, R. A. 2015. Variation in vowel duration among southern African American English speakers. *Am. J. Speech-Lang. Pat.* 24, 460-469.
- [8] Holt, Y., Jacewicz, E., Fox, R. A. 2016. Temporal variation in African American English: the distinctive use of vowel duration. *Journal of Phonetics & Audiology* 2:2 DOI: 10.4172/2471-9455.1000121.
- [9] Holt, Y. F., Rangaratham, B. 2018. F0 declination and reset in read speech of African American and White American women. *Speech Communication* 97, 43-50.
- [10] 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, 149–157.
- [11] 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.
- [12] IBM SPSS Statistics, v. 25. 2017. International Business Machines Corp. Armonk, NY.
- [13] Jacewicz, E., Fox, R. A. 2015. Intrinsic fundamental frequency of vowels is moderated by regional dialect. *J. Acoust. Soc. Am.* 138, EL405-EL410.
- [14] Jacewicz, E., Fox, R. A. 2018. Regional variation in fundamental frequency of American English vowels. *Phonetica* 75, 273-309.
- [15] Jacewicz, E., Fox, R. A., Wei, L. 2010. Between-speaker and within-speaker variation in speech tempo of American English. *J. Acoust. Soc. Am.* 128, 839-850.
- [16] Keating, P., Kuo, G. 2012. Comparison of speaking fundamental frequency in English and Mandarin. *J. Acoust. Soc. Am.* 132, 1050-1060.
- [17] Kingston J. 1992. The phonetics and phonology of perceptually motivated articulatory covariation. *Lang. Speech* 35, 99-113.
- [18] 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, 171–210.
- [19] Kingston, J., Diehl, R. 1994. Phonetic knowledge. *Language* 70, 419–454.
- [20] Labov, W., Ash, S., Boberg, C. 2006. *Atlas of North American English: Phonetics, Phonology, and Sound Change.* New York: Mouton de Gruyter.
- [21] Ladd, D. R., Silverman, K. E. A. 1984. Vowel intrinsic pitch in connected speech English. *Phonetica* 41, 31-40.
- [22] Lanehart, S. (ed). 2015. *The Oxford Handbook of African American Language.* Oxford: Oxford University Press.
- [23] Majewski, W., Hollien, H., Zalewski, J. 1972. Speaking fundamental frequency of Polish adult males. *Phonetica* 25, 119-125.
- [24] MathWorks Inc, Natick, MA. 2018. MATLAB R2018a. <http://www.mathworks.com>
- [25] Mennen, I., Schaeffler, F., Docherty, G. 2012. Cross-language differences in fundamental frequency range: A comparison of English and German. *J. Acoust. Soc. Am.* 131, 2249–2260.
- [26] Milenkovic, P. 2003. TF32 software program. Madison: University of Wisconsin.
- [27] Ohala, J. L., Eukel, B. W. 1987. Explaining the intrinsic pitch of vowels. In: Channon, R., Shockey, L. (eds). *In Honor of Ilse Lehiste.* Dordrecht: Foris Publications, 207-215.
- [28] Rickford, J. R. 1999. *African American Vernacular English: Features, Evolution, Educational Implications.* Oxford: Blackwell.
- [29] Scicon R and D Inc. 2005 PitchWorks.
- [30] Thomas, E. 2007. Phonological and phonetic characteristics of African American vernacular English. *Lang. Linguist. Compass* 1, 450-475.
- [31] Titze, I., R. 2000. *Principles of Voice Production, 2nd ed.* Iowa City: National Center for Voice and Speech.
- [32] Van Hoof, S., Verhoeven, J. 2011. Intrinsic vowel F0, the size of vowel inventories and second language acquisition. *J. Phonetics* 39, 168-177.
- [33] Whalen, D. H., Gick, B., Kumada, M., Honda, K. 1998. Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *J. Phonetics* 27, 125–142.
- [34] Whalen, D. H., Levitt, A. G. 1995. The universality of intrinsic F0 of vowels. *J. Phonetics* 23, 349-366.
- [35] Wolfram, W., Thomas, E. 2002. *The Development of African American English.* Oxford: Blackwell.