

ELICITING SOCIOPHONETIC VARIATION IN VOWEL DURATION

Ewa Jacewicz & Robert Allen Fox

Department of Speech and Hearing Science, The Ohio State University, USA
jacewicz.1@osu.edu, fox.2@osu.edu

ABSTRACT

Recent work has revealed that regional dialect is an important source of systematic variation in vowel duration in American English in addition to the well-known linguistic sources of this variation. This study addressed the issue of adequacy of speech materials used to elicit the dialect-specific temporal patterns. Duration of vowels was measured in isolated citation-form syllables and in connected speech of the same male and female talkers representing three different American English dialects. Despite the differences in absolute duration, the pattern of duration differences for the selected vowels was similar for each dialect and gender across the two types of production. The study provides evidence that the high level of control does not obscure dialect- and gender-specific patterns. It is concluded that citation-form vowels produced in isolation (representing fine control over phonetic context) can be successfully utilized to obtain valid samples of sociophonetic variation in vowel duration.

Keywords: Sociophonetics, vowel duration, dialect, gender, American English.

1. INTRODUCTION

It has been long established that vowels in American English (AE) vary in their duration in systematic ways. For example, vowels are shorter when followed by a voiceless consonant and longer when followed by a voiced consonant [1]. High vowels tend to be inherently shorter than low vowels and stressed vowels are longer than unstressed [2]. In connected speech, duration of vowels is further moderated by linguistic and paralinguistic factors related to discourse structure and sentence prosody including word- and phrase-final lengthening, semantic emphasis (focus), variations in lexical stress or emotional and physical state of the talkers along with their habitual speaking rate [3].

More recently, interest in vowel duration has been revitalized in the context of social factors and their effects on the temporal characteristics of speech. Importantly, both absolute and normalized vowel duration was found to vary systematically as a function of regional dialect, demonstrating that AE vowels in the Southern regions in the United States

are significantly longer than vowels in the North, Midland or West [4, 5, 6]. Sociophonetic variation was also reported for the effects of gender, showing that vowels produced by women tend to be longer than those produced by men [4, 7, 8]. Vowel duration can also be influenced by talker ethnicity. In particular, when controlled for regional dialect, African-American speakers were found to produce longer vowels relative to White AE speakers [8]. Finally, the socio-indexical sources of variation in vowel duration may include speaker age. To date, particularly well documented are temporal changes across language development in young children, whose lengthened vowels become progressively shorter with age [9, 10].

Given recent experimental evidence, vowel duration can now also be considered a sociophonetic variable. However, the sociophonetic status inevitably invites the question of the adequacy of speech materials used to study segmental duration in a social context. Speech samples can be obtained under a variety of conditions, ranging from reading lists of nonsense syllables in the lab to spontaneous conversations in a naturalistic setting. While conversational speech is favoured in sociolinguistic work, control over phonetic context (in the use of citation-form syllables) is desirable from the acoustic phonetic perspective. Unavoidably, variability in absolute durations across studies is extensive and, in large part, reflects differences associated with the speech materials used. As might be expected, studies using connected and conversational speech reported shorter average durations than those which opted for citation-form syllables [2, 7]. However, abandoning fine control altogether seems premature as it remains to be shown that there is indeed a significant difference between the *sociophonetic variation* in duration of vowels in connected speech and those produced in isolated citation-form syllables.

Addressing this issue, the current study examined the relevance of citation-form syllables to probe regional variation in vowel duration. It sought to determine whether the dialect-specific temporal patterns are independent of speech materials used and whether citation-form vowels can be utilized as a means to obtain valid samples of temporal behaviour of speakers of different dialects. To that end, vowel duration was assessed in two different

types of production by the same speakers, including citation-form syllables and connected speech.

2. METHODS

2.1. Participants

A total of 53 talkers ranging in age from 20-32 years participated. They represented three dialect regions in the US: the North, the Midlands and the South. The 18 Northern speakers (9 male, 9 female) were native of south-central Wisconsin, the 18 Midlands speakers (9 male, 9 female) came from central Ohio and the 17 Southern speakers (8 male, 9 female) were born and raised in western North Carolina. All participants were either college students or college graduates and had lived in their respective dialect areas for most of their lives.

2.2. Materials, procedures and measurements

The data presented here come from a larger study investigating regional dialect variation in AE across several generations of speakers [11, 12]. For the current analysis of vowel duration, five vowels were selected from a larger set of recordings: /ɪ, ε, e, æ, aɪ/. According to [2], these vowel categories include inherently short (/ɪ, ε/) and long (/e, æ/) vowels; the diphthong /aɪ/ represents the longest vocalic category. To elicit the citation-form syllables, the vowels were produced in the commonly used hVd-frame, yielding the prompts *hid*, *head*, *hayed*, *had* and *hide*. Three repetitions of each item were utilized in the current study for a total of 795 items from all 53 speakers. Each token appeared randomly as an individual prompt on the computer monitor and was recorded in isolation. A custom MATLAB program was used to control the experiment. The recordings were done either in a sound-attenuated booth in the lab or in a quiet room if a booth was unavailable. The participant spoke into a head-mounted Shure SM10A unidirectional microphone positioned 1.5 inches from his/her lips. The tokens were recorded and digitized at a 44.1-kHz sampling rate directly into a hard disc drive.

To elicit connected speech, each of the same vowel categories was produced in a unique sentence. The vowel was contained in the bVd-frame so that the monosyllabic target items were *bids*, *beds*, *bades*, *bads* and *bides*. Each target item carried the main sentence stress (and was the semantic focus), and was obtained in a contrastive stress paradigm as in the following example: *Ted thinks the fall SALES are low. No! Ted thinks the fall **BIDS** are low.* Each participant produced four repetitions of each sentence pair in a random order for a total of 1060 bVd-tokens from all 53 speakers. A different

MATLAB program was used to control the experiment, otherwise the procedure was as detailed above. Prior to recording each sentence set, the participant read it silently on the computer screen and was asked to produce it as naturally as possible (“the way you would say it if you were to contradict someone viciously”). A short familiarization practice was run prior to the actual experiment.

Prior to acoustic analysis, all tokens were digitally filtered and downsampled to 11.025 kHz. Vowel onsets and offsets were located by hand primarily on the basis of a waveform. Vowel onset was measured from onset of periodicity (at a zero crossing). Vowel offset was the point when the amplitude dropped to near zero. Reliability checks were done on all segmentation decisions using a custom program in MATLAB which displayed the landmark locations as vertical marks superimposed over a display of the waveform.

2.3. Predicted patterns

The emphatic vowels in the bVd-frame are expected to be lengthened as a function of the main sentence stress while reflecting the dialect-specific traits of conveying this emphasis. It is further assumed that individual vowel categories will maintain their relative durations independent of stress [2] but these intrinsic absolute durations will vary as a function of dialect, with NC vowels being the longest followed by OH and WI, respectively [4, 13]. Of interest here is whether, for each dialect, the durational patterns in the emphatic vowels correspond to those in the citation-form tokens, which are also lengthened due to their careful articulation when produced in isolation. Three possibilities arise: (1) both the citation-form vowels and the emphatic vowels are of comparable duration, (2) the emphatic vowels are significantly longer than the citation-form vowels, or (3) the citation-form vowels are significantly longer than the emphatic vowels. However, of greatest importance to the current investigation is whether, despite the differences in absolute duration, the pattern of duration differences across the five vowel categories is similar for each dialect across the types of speech materials used.

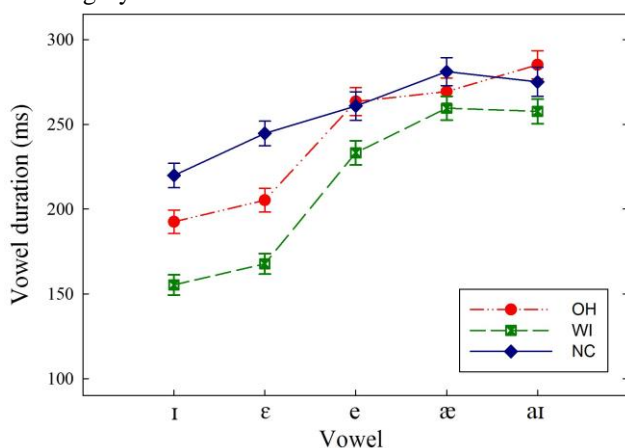
3. RESULTS

Data were analysed using a repeated-measures ANOVA with the within-subject factors vowel and stimulus type (citation-form, emphatic), and the between-subject factors dialect and gender. Where applicable, post hoc multiple comparisons were further made to explore a significant main effect or interaction.

3.1. The effects of vowel, dialect and gender

The main effect of vowel was significant [$F(4,188)=488.64$, $p<0.001$] showing that the duration of vowels increased with a greater degree of openness: /i, ε, e, æ, ai/, in that order. However, the main effect was weakened by a significant vowel by dialect interaction [$F(8,188)=23.58$, $p<0.001$], which arose because this general pattern was altered by dialect variation. In particular, the difference between the short and long vowels was larger in both OH and WI than in NC, where the short vowels /i, ε/ were comparatively longer. In general, the distinction between short and long vowels appeared to be minimized in NC, which is evident in Fig. 1.

Figure 1: Mean duration (s. e.) for each vowel category for each dialect.



The main effect of dialect was also significant [$F(2,47)=8.43$, $p=0.001$] and post hoc comparisons showed that WI vowels were significantly shorter than either OH or NC vowels, which did not differ significantly from one another. On average, NC vowels were the longest ($M=257$ ms) followed by OH ($M=243$ ms) and WI ($M=215$ ms), respectively. The significant main effect of gender [$F(1,47)=7.84$, $p=0.007$] indicated that female vowels were longer than male vowels. However, a significant dialect by gender interaction [$F(2,47)=4.06$, $p=0.024$] revealed that this was the case for both OH and NC but not for WI, where no significant differences due to talker gender were found.

3.2. The effects of stimulus type

The main effect of stimulus type was significant [$F(1,47)=8.49$, $p=0.005$] showing that the citation-form vowels were longer ($M=248$ ms) than the emphatic vowels ($M=228$ ms). However, a significant interaction between stimulus type and dialect [$F(2,47)=3.21$, $p=0.049$] arose because this

overall pattern was affected by dialect variation. This interaction is displayed in Figures 2-4.

Figure 2: Mean durations (s. e.) for WI vowels as a function of stimulus type.

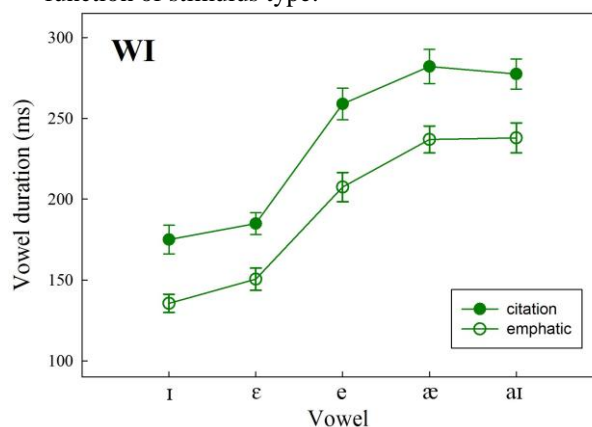


Figure 3: Mean durations (s. e.) for OH vowels as a function of stimulus type.

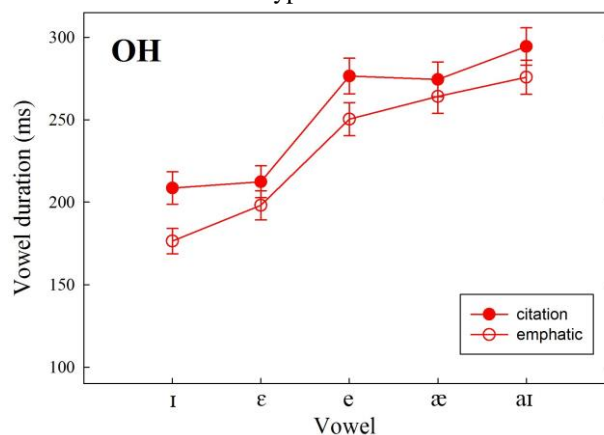
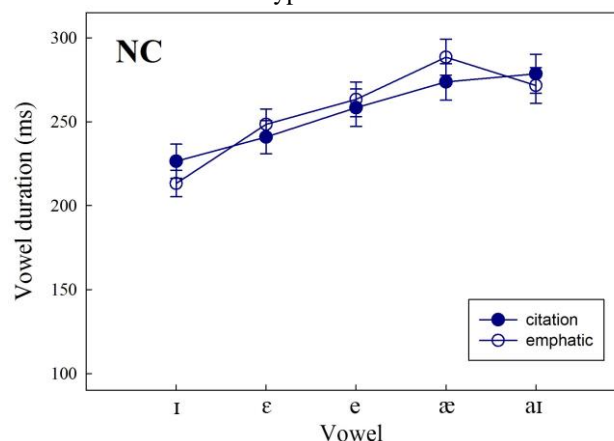


Figure 4: Mean durations (s. e.) for NC vowels as a function of stimulus type.



As can be seen, the effects of stimulus type were more variable across the dialects. For WI, the citation-form vowels were uniformly and considerably longer than the emphatic vowels and, despite the differences in absolute durations, the pattern of differences across all five vowels was

very similar. The citation-form vowels were also somewhat longer than the emphatic vowels for OH, although the differences between them were considerably smaller than for WI. One can also observe that the vowels /ε, æ/ in *head* and *had* approximated the durations of those in *beds* and *bad*s. For NC, the duration differences between the two types of production were negligible and, for three out of five vowels, the emphatic durations were even slightly greater than the citation-form durations. Post hoc tests showed that the negligible difference between the citation-form and emphatic productions in NC was significantly smaller than that in WI. The WI difference was numerically larger than the OH difference but did not represent a statistically significant difference. Not surprisingly, a 3-way interaction between vowel, stimulus type and dialect was also significant [$F(8,188)=2.54$, $p=0.012$], indicating that duration differences between the two types of productions varied with vowel category and were also influenced by dialect.

4. DISCUSSION

Several important findings emerged. Indisputably, the relative durations of individual vowels were maintained in each production type, indicating that speakers have knowledge of temporal relations holding among vowels in their own language. This knowledge is further shaped by their experience with regional variation which, in turn, determines the temporal characteristics of their speech to reflect the spoken language norms pertinent to a given geographic area. This acquired knowledge is necessarily present in their speech patterns and is thus independent of speech elicitation methods.

Furthermore, although citation-form syllables are produced in isolation, vowels in these syllables are not unnaturally lengthened. As found in this study, the citation-form vowels may be somewhat longer than the emphatic vowels produced to convey the main stress (and focus) of a sentence in read speech—for example, when reading a story to a child—but they are not always longer. They may also not differ substantially from the emphatic vowels, or may even be slightly shorter. This conclusion can be drawn on the basis of regional variation observed here. In particular, while the mean durational difference between the two types of production was 42 ms for WI, it was only 20 ms for OH and 2 ms for NC. In the latter case, the citation-form vowels were on average 2 ms shorter (and not longer) than the emphatic vowels.

These results underscore the importance of regional variation in understanding the patterns of segmental durations. Clearly, the NC vowels were

the longest from all three dialects and the contrast between the inherently short and long vowels was minimized. These temporal relations among vowels seem to represent a distinct property of Southern AE where the durational contrast between tense and lax vowels may be reduced relative to other dialects in the US [6]. Importantly, these temporal relations were “preserved” in citation-form vowels.

Regional variation in vowel duration and segmental timing in general have not been widely studied and relatively little is known about temporal patterns across regions in the US, although recent research has begun to explore aspects of temporal organization in AE [6, 14]. The current study contributes further evidence that dialects differ in the way the segmental durations are utilized and that the choice of speech materials does not obscure the dialect-specific patterns. The current data are also consistent with previous findings that WI vowels are shorter than OH and NC vowels, respectively, and that these regional differences are manifested in speakers of different ages: they are evident in young adults studied here, in older adults [15] and in children [16]. However, it needs to be emphasized that dialect differences may not be manifested across all vowel categories all the time. In general, more work is needed to better understand the regional variation in segmental timing.

The current study also found gender-related differences, with female vowels being significantly longer than male vowels. This finding is in line with several earlier reports including [7], who found statistical differences between longer female vowels and shorter male vowels on the basis of citation-form hVd-syllables produced by a large number of speakers from a common geographic area. Importantly, the effects of gender in the current study were independent of production type as none of the relevant interactions with gender (i.e., gender x stimulus type and gender x stimulus type x vowel) were significant. These results underscore the general conclusion stemming from this study that highly controlled laboratory conditions do not obscure differences in vowel duration related to speaker characteristics. As a contribution to the current debate about adequacy of speech materials for particular interpretations of phonetic data [17], this research found nothing problematic with using citation-form syllables to study vowel duration patterns across dialects, gender and age.

5. ACKNOWLEDGEMENTS

This work was supported by NIDCD Grant R01DC006871. We thank Joseph Salmons for his contributions to this research.

6. REFERENCES

- [1] Peterson, G. E., Lehiste, I. 1960. Duration of syllable nuclei in English. *J. Acoust. Soc. Am.* 32, 693-703.
- [2] Crystal, T. H., House, A. S. 1988. The duration of American-English vowels: An overview. *J. Phonetics* 16, 263-284.
- [3] Klatt, D. H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59, 1208-1221.
- [4] Jacewicz, E., Fox, R. A., Salmons, J. 2007. Vowel duration in three American English dialects. *Am. Speech* 84, 367-385.
- [5] Tauberer, J., Evanini, K. 2009. Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. *Proc. INTERSPEECH 2009* Brighton, 2211-2214.
- [6] Fridland, V., Kendall, T., Farrington, C. 2014. Durational and spectral differences in American English vowels: Dialect variation within and across regions. *J. Acoust. Soc. Am.* 136, 341-349.
- [7] Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099-3111.
- [8] Holt, Y. F., Jacewicz, E., Fox, R. A. (to appear). Variation in vowel duration among southern African American English speakers. *Am. J. Speech Lang. Pathol.*
- [9] Lee, S., Potamianos, A., Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 103, 1455-1468.
- [10] Ballard, K. J., Djaja, D., Arciuli, J., James, D., van Doorn, J. 2012. Developmental trajectory of production of prosody: Lexical stress contrastivity in children ages 3 to 7 years and in adults. *J. Speech Lang. Hear. Res.* 55, 1822-1835.
- [11] Jacewicz, E., Fox, R., A., Salmons, J. 2011. Cross-generational vowel change in American English. *Lang. Var. Change* 23, 45-86.
- [12] 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.
- [13] Jacewicz, E., Fox, R., A., Salmons, J. 2011. Vowel change across three age groups of speakers in three regional varieties of American English. *J. Phonetics* 39, 683-693.
- [14] Clopper, C. G., Smiljanic, R. 2015. Regional variation in temporal organization in American English. *J. Phonetics* 49, 1-15.
- [15] Fox, R. A., Jacewicz, E. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *J. Acoust. Soc. Am.* 126, 2603-2618.
- [16] Jacewicz, E., Fox, R., A., Salmons, J. 2011. Regional dialect variation in the vowel systems of typically developing children. *J. Speech Lang. Hear. Res.* 54, 448-470.
- [17] Wagner, P., Trouvain, J., Zimmerer, F. 2015. In defense of stylistic diversity in speech research. *J. Phonetics* 48, 1-12.