Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English

Joshua Tauberer, Keelan Evanini

Department of Linguistics, University of Pennsylvania, Philadelphia, USA
{tauberer,keelan2}@ling.upenn.edu

Abstract

We report the results of a comprehensive dialectal survey of three vowel duration phenomena in North American English: gross duration differences between dialects, the effect of post-vocalic consonant voicing, and intrinsic vowel duration. Duration data, from HMM-based forced alignment of phones in the Atlas of North American English corpus [1], showed that 1) the post-vocalic voicing effect appears in every dialect region and all but one dialect, and 2) dialectal variation in first formant frequency appears to be independent of intrinsic vowel duration. This second result adds evidence that intrinsic vowel durations are targets stored in the grammar and do not result from physiological constraints.

Index Terms: vowel duration, North American English dialects, consonant voicing, intrinsic duration

1. Introduction

We investigated two factors that influence vowel duration in North American English — post-vocalic consonant voicing and intrinsic vowel duration — across the dialects of the Atlas of North American English [1], henceforth ANAE. Our findings, the most comprehensive that we know of to date, contribute to the debate in the literature as to whether the two duration phenomena are merely physiological in nature or are part of the grammar. We also report overall vowel duration differences by dialect.

The duration of a vowel is longer when preceding a voiced versus voiceless obstruent (e.g. [2]), a so-called post-vocalic consonant voicing effect. The difference is quite robust in prepausal position, reported often as a ratio of voiced-to-voiceless duration around 1.4 [3]. The ratio decreases substantially and the effect may disappear in fluent speech. The duration difference is also larger for vowels with greater intrinsic duration [4]. The post-vocalic voicing effect has been reported in a number of languages but is not universal [5],[6]. Still, English shows a much larger duration difference between voiced and unvoiced consonants than all other languages in which the effect has been studied.

Though the voicing effect has been investigated specifically in several dialects of American English (Alabama, Chicago, Los Angeles Chicano, and Jamaican Creole [7]; Wisconsin, Ohio, and North Carolina [8]), we report the first comprehensive survey across North American English. If we find that the voicing effect differs in different dialects of English, it will contribute to our understanding of the grammatical status of the effect.

The second factor we observed is the variation that exists between vowels. [2],[9] observed that mean durations of vowels vary from vowel to vowel, on the order of 30ms, and that these means are correlated with vowel height: the lower the vowel the greater its "intrinsic" duration. Still, as [10] discussed, there are two types of explanations for this phenomenon. One, taken by [11], is that the differences are purely physiological: low vowels require greater jaw opening and so require more time to articulate. The other, taken by [10], is that each vowel has a duration target specified in the grammar. [10] noted, for instance, that low vowels were reported not to have longer transitions, as would be expected if the duration difference was due to jaw movement, but instead longer steady states. It is possible, then, that the correlation with vowel height is synchronically incidental, and current duration targets were grammaticized from an earlier purely physiological pattern.

More evidence for the linguistic (i.e. grammatical) nature of intrinsic vowel duration comes from apparent distinctive duration contrasts in two dialects of American English [12]. In Pittsburgh, AH and monophthongal AW overlap in formant space but differ by approximately 100ms in duration; in the Inland North EH and AA overlap in formant space but differ by 50ms. This appears to be a type of phonological length, that is, a duration difference unexplainable in physiological terms.

A dialectal study of intrinsic vowel duration leads to the question of what happens to intrinsic duration as a vowel undergoes sound change, in particular in terms of height. As a vowel is lowered, e.g., through a chain shift, the physiological explanation predicts the vowel's intrinsic duration will increase while the phonological explanation predicts that the duration target will remain unchanged — other things being equal. We test these hypotheses in Section 3.3.

2. Methods

2.1. Description of the ANAE corpus

2.1.1. Speakers and sampling methods

The ANAE was collected in an attempt to provide a comprehensive view of the current sound changes in progress in all of the major dialect regions of North America. Previous corpora that contained dialect variation were not adequate for dialectological purposes since they did not control well for the geographic background of the speakers and did not provide a broad sample of cities and subregions from within the major dialect regions.

The ANAE interviews were conducted over the telephone so that speakers from all regions could be accessed efficiently. The sampling methods ensured that the corpus contains more accurate and more fine-grained dialect information than existing corpora: at least two speakers were selected randomly from every city in North America with at least 50,000 inhabitants, and only speakers who had lived their entire lives in that city were chosen. In total, 762 speakers were interviewed for the ANAE. Of these, a subset of 439 were selected for detailed acoustic analysis by the ANAE authors. Interviews were a mix

of spontaneous speech, minimal pair tests, and other elicitation methods, though the speech style was not consistently recorded and so can not be controlled for below.

Table 1 provides the dialect region affiliation of the speakers in the portion of the corpus used in this study. In addition to dialect region, we also investigated phenomena at the more specific dialect level (as defined in [1]) and report results in Section 3.2 for the Boston and Maine speakers, both part of the Eastern New England dialect region.

Dialect Region	Speakers	Tokens
North	124	26,299
South (Region)	76	21,814
Midland	63	15,335
West	41	9,123
Canada	28	7,089
Western PA	13	3,685
Mid-Atlantic	12	2,950
Eastern New England (ENE)	10	2,141
Southeast	10	2,838
New York City (NYC)	5	1,706
Total	*382*	*92,980*
Dialect	Speakers	Tokens
Boston	5	925
Maine	2	497
Inland North	61	13,343
Pittsburgh	6	1,810
South (Dialect)	58	16,672

Table 1: Counts of speakers and tokens used in this study by dialect region and dialect in the ANAE corpus.

2.1.2. Duration measurements

In [1], F1 and F2 measurements were taken by hand for all vowels, with a focus on vowels that are undergoing change in each region. Formant measurements for each speaker were subsequently normalized using the procedure in [13]. In total, ca. 300 vowels (always bearing primary lexical and phrasal stress) were analyzed for each of the 439 speakers. Each measured token was extracted into a separate audio file — this collection of 133,723 individual word files comprises the audio database of the ANAE corpus.[1]

Because the ANAE is a database of sound change across dialects, vowel labels were assigned not by the perceived vowel quality (e.g. as an IPA symbol) but instead according to phonological classes. That is, the so-called short-o vowel, denoted AA in ARPABET, may have quite disparate phonetic realizations across dialects but are labeled all as AA in the corpus. In a sense, then, vowels are actually lexical classes. We limited our analysis to the 15 vowels (i.e. phonemes) indicated in Figure 1, which has an example word for each vowel (from J.C. Wells' lexical classes as reproduced in [1, p13]).[2]

In order to obtain duration measurements for the vowels in each of the tokens, the corpus was processed with the forced alignment system described in [14]. The system uses GMM-based monophone HMM acoustic models with 32 mixture components on 39 PLP coefficients trained on 25.5 hours of speech from the SCOTUS corpus.

2.2. Duration normalization

We normalized durations in Sections 3.2 and 3.3, not by speaker, as is usually done, but in order to minimize potentially confounding contextual effects not of interest to our study. A correlation in the corpus between vowel quality and the voicing of the following consonant, for example, would undermine the interpretation of the results; factoring out variables not of interest also reduces the variation in the remaining data. We normalized durations by fitting a linear model to the log-duration data and then subtracting from each duration measurement the predicted components due to the unwanted factors. In Section 3.2 on post-vocalic voicing, the model contained vowel class and post-vocalic place and manner of articulation. In Section 3.3 the model contained post-vocalic place and manner of articulation, voicing, and number of following syllables. (In Section 3.2 only word-final syllables were used.)

The log-duration model we chose treats each factor as having a multiplicative effect on duration, rather than an effect in absolute terms (i.e. seconds). It is somewhere between a simple linear model and the more complex model proposed by [15] based on the notion of "incompressibility" — that each successive shortening effect on a vowel has a diminished effect because of physical bounds on the speed of articulation. Because the duration phenomena do not combine precisely multiplicatively, some confounding correlations no doubt remain after this process.

3. Results

3.1. Vowel duration by region

Table 2 reports mean duration in word-final syllables by dialect region. The shortest region was New York City at 133ms, while the South and Southeast regions had the longest mean durations at 156ms and 159ms, respectively. The differences between the South and the two regions ranked directly below it, the Midland and the West, were not significant, but to the next region, Western PA, the difference was significant (Tukey post-hoc $p < .01$).

Region	Duration	Region	Duration
NYC	133	Western PA	150
ENE	140	West	153
Canada	142	Midland	154
Mid-Atlantic	146	South	156
North	149	Southeast	159

Table 2: Mean durations (ms) of vowels in word-final syllables by dialect region.

At first glance, Table 2 might seem to underlie the commonly held perception that Southerners speak with a slower overall speaking rate than other regions. However, we found no such regional difference in a large corpus of spontaneous speech containing regional variation [16]. The mean speaking rates (determined by excluding filled pauses and only considering utterances containing at least five words) for 445 Northern and 1,421 Southern speakers from the Fisher corpus are both 193 words per minute (it is impossible to calculate speaking rate

[1] We excluded the 26 speakers labeled as "T" in the ANAE corpus who do not form a coherent dialect region, and a small portion of the audio files were excluded due to poor quality. In total, the results presented are based on analysis of 92,980 tokens from 382 speakers.

[2] We retain the separation in the data of AY before voiceless consonants ("AY(0)") and all other occurrences of AY ("AY(V)").

in the ANAE corpus, since the interviews were not transcribed). Speaking rate, then, does not explain the vowel duration difference observed.

A more likely explanation for the fact that Southerners have a longer overall mean vowel duration is based on the nature of the Southern Shift, a large-scale vowel change currently in progress in the South (see Section 18.3 in [1] for a complete description of the shift). In this change, the lax vowels IH and EH become tense (more peripheral) and develop a schwa-like off-glide. The vowel-specific plots in Figure 1 show that the South is at or close to the top in the duration rankings for these vowels, as well as for the lax vowels AA and AH. This suggests that the lengthening of IH and EH as part of the Southern Shift may be causing a larger reanalysis of phonological duration for the lax vowels in the South. This interpretation corresponds well with the finding in [8] that IH, EH, and AE were longer in the South than the North or Midland, and the finding in [17] that IH, EH, AH, and UH were longer in the South than most other regions. Furthermore, the nuclei of IY and EY are lowered substantially, leading to a potential increase in intrinsic vowel duration (see Section 3.3). All of these factors may contribute to the fact that the overall mean vowel duration is longest in the South.

3.2. Post-vocalic consonant voicing

Table 3 reports the vowel duration ratio (mean duration before voiced obstruents to that before voiceless obstruents) in each of the dialect regions for vowels preceding a stop, fricative, or affricate. The dialect regions show similar duration ratios in the range of 1.13–1.27 corresponding to an 11–25ms difference. A duration ratio around 1.2 is what would be expected given the nature and mix of the speech tasks in the corpus. Durations were normalized as described above.

The only outliers among the dialects were the Boston and Maine dialects, with ratios of 1.33 and 1.02, respectively. (These dialects also had a very small number of tokens applicable for analysis in this section — 356 and 175 — from just six and two speakers, respectively.) The Maine dialect's duration ratio at 1.02 is considerably less than what has been found in any comparable study of English, but a regression analysis of normalized log-durations showed the interaction between voicing and membership in the Maine dialect to be nonsignificant. The interaction for the Boston dialect was significant ($p < .02$).

Region	Ratio	Region	Ratio
NYC	1.13	North	1.23
South	1.16	ENE	1.24
Canada	1.19	Southeast	1.24
West	1.19	Mid-Atlantic	1.25
Midland	1.21	Western PA	1.27
Dialect	Ratio	Dialect	Ratio
Maine	1.02	Boston	1.33

Table 3: Post-vocalic voicing duration effect as a ratio (pre-voiced vowel duration to pre-voiceless duration) for the dialect regions and two dialects.

Some of the dialect differences can be attributable to overall vowel duration differences. The dialect region with the shortest vowels, NYC, also had the smallest voiced–unvoiced ratio, as expected based on incompressibility. On the other hand, the South, with some of the longest vowels, had a short duration ratio as well.

3.3. Intrinsic vowel duration

It has long been observed that a vowel's height is related to its duration: low vowels (i.e. high F1) tend to have longer intrinsic duration. Our findings from the ANAE corpus also demonstrate this. Across the 15 vowel classes, the correlation between mean F1 and mean duration is strong (r=.68); the slope of the regression line is 18ms of duration per 100 Hz increase in F1. Both a physiological and phonological explanation of intrinsic vowel duration predict this result.

Our hypothesis was that if the physiological explanation were true, then as a vowel class changes in F1 due to regular sound change it should also change in mean duration (a positive F1–duration correlation). If the phonological explanation were true, the mean duration would stay the same. To test this hypothesis, we considered each vowel class one at a time. For each, we computed the correlation between mean F1 and duration across dialects. For this analysis only vowels preceding obstruents were considered so that the geographically widespread AE-tensing before nasals did not obscure the unique dialectal difference of this vowel in the Inland North, which has AE-tensing everywhere.

Just two vowel classes had statistically significant correlations ($\alpha = .05$): IH and EH, both with negative correlations. Scatter plots are shown in Figure 1. None of the vowels showed the positive correlation that would be expected under the physiological model of intrinsic duration. Nevertheless, correlations are difficult to interpret here. Not only do the nuclear targets of the vowels change across dialects, but so do their trajectories. Monophthongization and diphthongization processes might increase or decrease intrinsic vowel duration for reasons independent of vowel height, though this has yet to be shown empirically.

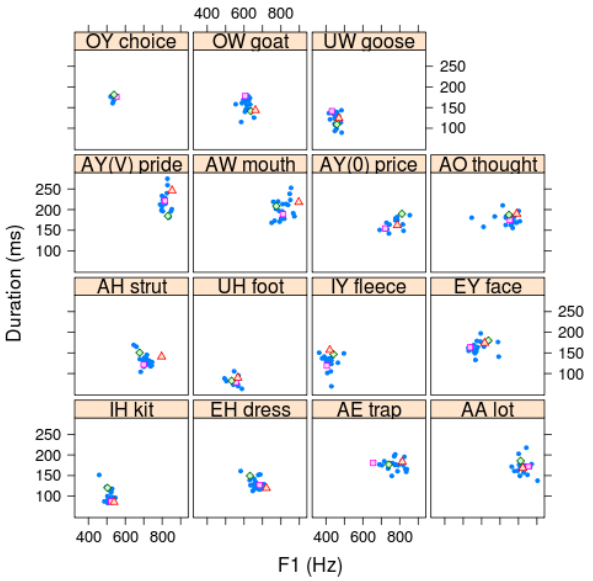

Figure 1: Mean vowel duration against mean first formant frequency, by vowel. Points are dialects. Key: Inland North–purple square, South–green diamond, Pittsburgh–red triangle

A specific regional sound change that provides a clear test for the physiological and phonological explanations of intrinsic duration is the lowering of AH as part of the Pittsburgh shift (see Section 19.3 in [1] for a description of the changes involved). The group F1 mean for Pittsburgh AH, 785 Hz, is 50–100 Hz higher than for all other dialects (clearly visible in Figure 1), and AH overlaps considerably in F1 space with monophthongized AW. A pure physiological explanation for the overall effect of F1 on duration would suggest that the lowering of AH in Pittsburgh would lead to it also having a longer duration in this dialect. As shown in Figure 1, however, Pittsburgh AH is clearly located in the middle of the distribution of durations across dialects. This would tend to support the explanation that intrinsic duration is largely phonological.

Another informative vowel is AE. Figure 1 shows that the F1 variation by dialect is quite large; situated at the extreme low end of F1 is the group mean for the Inland North speakers. Stage 1 of the Northern Cities Shift, which operates nearly uniformly across the Inland North, involves across-the-board raising of AE. Other dialects have raising in more restricted environments: before /d/ and /g/ for many Midland speakers, before voiced obstruents and voiceless fricatives in NYC (see Section 13.2 in [1]), and before nasals in all dialects except Canada. For this reason tokens before nasals were excluded from our data, since they would obscure the dialect-specific patterns for AE. Despite this widespread variation in F1 for AE, the distribution of mean duration values across dialects is quite uniform, and the region with the most raising, the Inland North, is clearly in the middle. Again, as with AH in Pittsburgh, this would seem to indicate that AE has a phoneme-specific duration target that is not influenced by regional differences in F1. This conclusion is not quite as clear as in the case of AH, though, since many dialects with raised AE also develop a schwa-like off-glide which could lead to increased duration and obscure the change in duration due to a physiological effect based on height. Further research into the relationship between duration and vowel formant trajectories would help clarify the case of AE.

4. Conclusions

Automated forced alignment has proven to be a useful tool for sociophonetic research. Applied to the ANAE corpus, duration measurements for over 100,000 tokens spanning the major dialects of American English have been readily obtained, without any significant manual labor. Though any individual phone boundary has a fair amount of error, mean durations appear to be reliable due to the law of large numbers.

We reported three major findings. The first was that vowel durations in South speech are greater than elsewhere, despite no similar difference at the level of speaking rate, and we noted why South speech might differ in this way from other dialect regions. This replicated previous reports.

The post-vocalic consonant voicing effect on vowel duration is a curious phenomenon — while it exists in many languages, English exhibits a particularly large duration difference. These differences have most often been examined in lab speech either of a single dialect or of unknown dialect, though we noted several dialectal studies in the introduction. With our corpus we were able to show that the voicing effect exists throughout the dialects of North American English, all with very similar duration differences. The most notable exception was the two speakers from Maine who showed only a marginal voicing effect, though the size of this sample indicates that further research in this region is warranted.

The last question we addressed was whether intrinsic vowel duration could be accounted for in terms of vowel height. Dialectal variation brings new evidence to this debate. We failed to find any indication that duration is affected by sound change. As a vowel's height increases, a physiological explanation of intrinsic duration predicts duration to decrease. If anything, we found a tendency for duration to increase. In the case of AE especially, which has a wide F1 spread across dialects with no apparent correlation to duration, a phonological or linguistic model of intrinsic duration in terms of a duration target seems most likely. All this being said, variation in vowel formant trajectories between dialects, which is not reflected in the static F1 measurements from the corpus, is a potential confounding factor yet to be explored.

5. Acknowledgments

We would like to thank Bill Labov, Mark Liberman, Jiahong Yuan, Steve Isard, and the remainder of our phonetics lab Splunch group.

6. References

[1] W. Labov, S. Ash, and C. Boberg, *The Atlas of North American English*. Mouton de Gruyter, 2006.

[2] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *J. Acoust. Soc. Am.*, vol. 25, pp. 105–113, 1953.

[3] N. Umeda, "Vowel duration in American English," *J. Acoust. Soc. Am.*, vol. 58, no. 2, pp. 434–445, 1975.

[4] P. A. Luce and J. Charles-Luce, "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," *J. Acoust. Soc. Am.*, vol. 78, pp. 1949–1957, 1985.

[5] K. R. Kluender, R. L. Diehl, and B. A. Wright, "Vowel-length differences before voiced and voiceless consonants: an auditory explanation," *Journal of Phonetics*, vol. 16, pp. 153–169, 1988.

[6] K. de Jong and B. Zawaydeh, "Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration," *Journal of Phonetics*, vol. 30, pp. 53–75, 2002.

[7] T. C. Veatch, "English vowels: Their surface phonology and phonetic implementation in vernacular dialects," Ph.D. dissertation, University of Pennsylvania, 1991.

[8] E. Jacewicz, R. A. Fox, and J. Salmons, "Vowel duration in three American English dialects," *American Speech*, vol. 82, no. 4, pp. 367–385, 2007.

[9] G. E. Peterson and I. Lehiste, "Duration of syllable nuclei in english," *The Journal of the Acoustical Society of America*, vol. 32, no. 6, pp. 693–703, 1960.

[10] L. Lisker, "On 'explaining' vowel duration variation," *Glossa*, vol. 8, no. 2, pp. 233–246, 1974.

[11] I. Lehiste, *Suprasegmentals*. Cambridge: MIT Press, 1970.

[12] W. Labov and M. Baranowski, "50 msec," *Language Variation and Change*, vol. 18, pp. 223–240, 2006.

[13] T. Nearey, "Phonetic feature system for vowels," Ph.D. dissertation, University of Connecticut, 1977.

[14] J. Yuan and M. Liberman, "Speaker identification on the SCOTUS corpus," in *Proceedings of Acoustics '08*, 2008.

[15] D. H. Klatt, "Interaction between two factors that influence vowel duration," *J. Acoust. Soc. Am.*, vol. 54, no. 4, pp. 1102–1104, 1973.

[16] C. Cieri, D. Miller, and K. Walker, "The Fisher corpus: a resource for the next generations of speech-to-text," 2004.

[17] C. G. Clopper, D. B. Pisoni, and K. de Jong, "Acoustic characteristics of the vowel systems of six regional varieties of American English," *Journal of the Acoustical Society of America*, vol. 118, no. 3, pp. 1661–1676, 2005.