ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Predictability affects vowel dispersion and dynamics in the Buckeye Corpus

Michael McAuliffe & Molly Babel

Department of Linguistics, University of British Columbia, Vancouver, BC, Canada
mcauliff@interchange.ubc.ca, molly.babel@ubc.ca

Abstract

Research on vowel space dispersion has found that vowel spaces are less dispersed when lexical and contextual factors favor identification of a word (e. g., a highly predictable word), and vowel spaces are more dispersed when lexical and contextual factors create an environment for less predictability [1, 2]. Examining sound changes in progress, recent work has demonstrated that, for some vowel changes, a vowel will be produced in a more innovative way when the lexical item in which it occurs is highly semantically predictable [3]. The data for such claims have been collected in laboratory settings. In this paper we investigate whether these findings from the laboratory extend into more naturalistic settings, and we examine the precise character of the reduction by examining dynamic movement across a vowel's duration.

Index Terms: reduction in speech production, spontaneous speech, acoustics, corpus linguistics, SSANOVA

1. Introduction

Vowels in predictable words tend to be reduced relative to their unpredictable counterparts. Specifically, more predictable words, such as those with a high frequency of occurrence or those in sparse phonological neighborhoods, have vowels that are typically produced in a more reduced acoustic-phonetic space compared to less predictable words [1, 2]. Vowels in more predictable words also tend to show more advanced realizations of a sound change [3], though not every sound change manifests this way. The research focusing on these aspects has used single-word or single-utterance productions rather than natural conversation. Laboratory-based speech is generally more clear and characterized by expanded vowel spaces, as compared to conversational speech styles [4]. In this paper, we investigate whether the factors that have been found to influence vowel spaces in citation styles also influence vowel spaces in more natural contexts.

Perhaps more importantly, past research on vowel space dispersion has used single point measurements of vowels. However, the dynamic nature of vowels is well known, e. g. [5, 6], and recent research has moved towards incorporating measurements of multiple points along a vowel into speech research, e. g. [7, 8]. In this paper we use a statistical technique known as smoothing spline ANOVA (henceforth, SSANOVA). SSANOVA is a model that compares shapes of curves to determine if they are significantly different from one another and exactly where those differences lie [9]. The technique has been used in linguistics for use with ultrasound images of tongues [10], and has recently expanded to use with formant measurements [11, 12, 13].

The goals of this paper are, therefore, two-fold. First, we seek to determine whether the previously documented patterns of vowel reduction as a function of predictability from the lab extend to spontaneous speech. The second goal is to determine whether vowel dynamics provide a different perspective on reduction.

2. Methodology

2.1. Materials

The Buckeye Corpus [14] is an ideal place to test whether the findings of previous research hold outside of an laboratory setting. The corpus was recorded and annotated for the purposes of acoustic analysis. Forty speakers from Columbus, OH participated in the creation of the corpus. The Buckeye Corpus classifies speakers according to age, with those over 40 in the old group, and those under 30 in the young group. For the current study, only the 20 female speakers in the corpus were analyzed to avoid any gender-based variation in vowel spaces [15].

2.1.1. Data selection

Data for the current study were extracted using the transcriptions and sound files of the corpus. The words to be analyzed were ones that (1) did not contain any nasal segments in their transcription, (2) had /æ, ɪ, i, u, ʊ, ɑ/ as the stressed vowel, (3) had their stressed vowel surrounded by obstruents, (4) were monosyllabic, and (5) were content words. These criteria were chosen in order to match the data in this study as closely as possible to those in previous studies [1, 2]. For each of the words that met these criteria, the start and end points of the stressed vowel were used to measure the spectral characteristics in Praat [16]. A total of 2,564 vowels were analyzed (/æ/: 514, /ɪ/: 739, /i/: 214, /u/: 67, /ʊ/: 443, /ɑ/: 587), with 1,491 vowels taken from older speakers and 1,073 vowels taken from younger speakers.

2.2. Factors considered

The lexical factors that are examined in this study are frequency and neighborhood density. Frequency of a word was calculated by dividing the number of occurrences within the corpus over all 40 speakers by the total number of words uttered by all 40 speakers and then log transformed [17]. The frequencies calculated were significantly correlated with frequencies from larger corpora, such as SUBTLEXus [18] ($r = 0.85$, $t(2098) = 75.5$, $p < 0.01$). Neighborhood density is the count of all neighbors of a word, where a neighbor is defined as a word that differs from the given word by a single insertion, deletion or substitution at any point in the transcription [19].

A key difference between conversational speech and words in isolation is that the context for a word is vastly different for each. Research on the realization of words in spontaneous speech corpora has focused on duration measurements, e. g. [17, 20], but we include the factors that have been found to influence duration to control for possible effects of contextual information. Conditional probability and the number of repetitions were calculated. The conditional probability of a word given the preceding word and the conditional probability given the following word were calculated by dividing the total number of times that the two words appeared together by the total number of times that the preceding/following word appeared in the corpus [17]. Repetition was calculated by counting the number of times that a word had appeared in the interview up to that point, though values above 20 were grouped together into one level. Both conditional probabilities and repetition were log transformed [17].

2.3. Analysis

Formants from the stressed vowels were extracted using 10th-order Linear Predictive Coding [16], though for some speakers higher orders were used if formant tracking was inconsistent. The first and second formants over the course of the vowel were compiled and analyzed in `R` [21]. Time values were converted into percentage of duration, and the first and second formants were vowel-extrinsically normalized to represent a percentage of a speaker's range for that formant to account for individual differences in vowel production [22].

For comparison to earlier research [2], values from the midpoint of each vowel were extracted. The normalized formant values at the midpoints were then analyzed with a linear regression model with relative frequency and neighborhood density as continuous variables in the model. The control variables that were included were the previous and following conditional probability and repetition, which have been found to affect word duration in previous studies [17]. These results are reported in Section 3.1. To gain a better understanding of lexical factors' effect on vowel dynamics, we used all measurements through a vowel to create SSANOVA models using the `gss` package [23]. This analysis used normalized time (percentage of total duration) and normalized F1 and F2 and is presented in Section 3.2.

3. Results

3.1. Midpoint analysis

A factor was deemed significant if its inclusion significantly improved the fit of the model [17]. To generate the following graphs, fitted models were used to predict F1 and F2 values at the first and third quartiles of a given distribution of a continuous factor. Larger point size in the graphs corresponds to vowels from low frequency words and words in dense phonological neighborhoods. Following previous findings, we predict that these vowels would occupy more of the periphery of the vowel space.

Frequency was not significant in the model of F1 ($F(1, 2528) = 1.57$, $p = 0.18$), but was a significant contributor in the model of F2 ($F(1, 2528) = 16.13$, $p < 0.05$) The additional variance accounted for when frequency was added to the model (ΔR^2) was 1.2%, which is similar to the ΔR^2 for word frequency in models of duration in spontaneous speech [17]. As can be seen in Figure 1, the vowels that showed the most centralization in high frequency words were /ʊ/ and /ɑ/. The inconsistent behavior of /u/ is likely due to the relative rarity of monosyllabic words with /u/ in the dataset (n = 67).

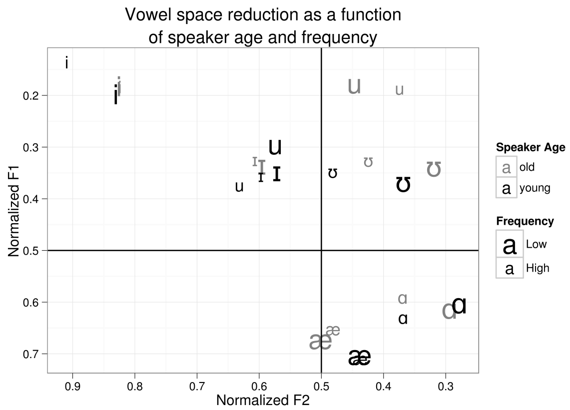

Figure 1: *Effect of word frequency on midpoints of vowel realizations by older and younger speakers.*

Like word frequency, neighborhood density was also not a significant contributor in the model of F1 ($F(1, 2528) = 1.45$, $p = 0.52$), but was a significant contributor in the model of F2 ($F(1, 2528) = 5.67$, $p < 0.05$). The additional variance accounted for in the model with neighborhood density as a predictor (ΔR^2) was 0.6%, which is half the additional variance accounted for when frequency was added to the model. In Figure 2, the vowels that showed the most centralization in sparse neighborhoods were /u/ and /ʊ/, but they also showed generational effects; young speakers have an advanced /u/ (higher F2) in all contexts, but older speakers only have advanced /u/ in words in sparse neighborhoods. For /ʊ/, younger speakers have a more advanced /ʊ/ in words in sparse neighborhoods than in words in dense neighborhoods, but older speakers are more consistent.

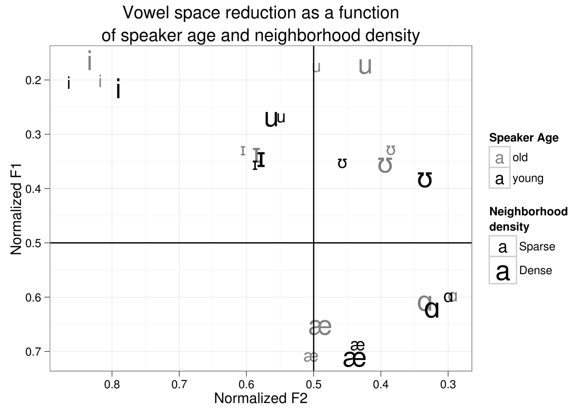

Figure 2: *Effect of neighborhood density on midpoints of vowel realizations by older and younger speakers.*

The lexical factors examined here significantly contributed to the linear regression models of F2 for these speakers, but not to models of F1. In addition, frequency accounted for more additional variance in predicting F2 than neighborhood density. Generational effects can also be seen. Younger speakers appear to produce /u/ farther forward in the vowel space than older speakers, and /æ/ farther back in the vowel space than older speakers.

3.2. SSANOVA analysis

The following graphs were generated in a similar way to the graphs above, but instead of linear regression, SSANOVA models were used instead to predict F1 and F2 to generate movements through formant space from 20% to 80% of total duration. The same method of using the first and third quartiles of a continuous distribution was used. Significance for SSANOVA models is generally assessed using visual inspection of 95% confidence intervals around the splines [10, 11]. However, given the complex nature of plotting a spline against another one, we use this technique in a more exploratory and qualitative approach to vowel trajectories. Figure 3 shows the differences in formant trajectories between vowels in high and low frequency words by the two speaker groups. Like in the midpoint analysis, we see that /ɑ/ and /ʊ/ are the most centralized, and their dynamics are the most severely affected by lexical frequency. For older speakers, /i/ also becomes more centralized in high frequency words. Older speakers use a more retracted /u/ in low frequency words than in high frequency words.

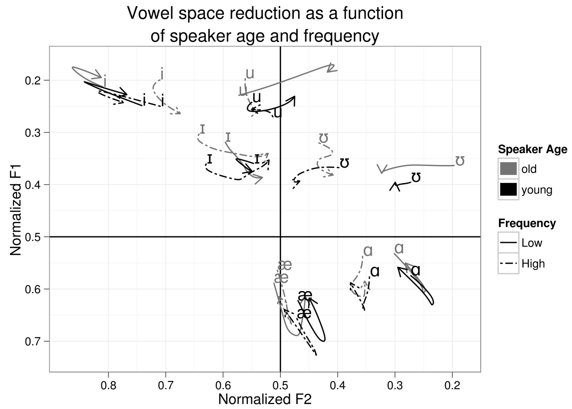

Figure 3: *Effect of word frequency on trajectories of vowel realizations by older and younger speakers.*

Figure 4 shows the differences in formant trajectories between vowels in words in sparse and dense neighborhoods by the two speaker groups. It is less clear in this graph whether neighborhood density affects vowel trajectories in a systematic way. From previous research, we expect that vowels from words in dense neighborhoods should be produced more clearly than from words in sparse neighborhoods. However, in most cases, vowels from words in dense neighborhoods are more condensed, such as with /ɑ/ and /æ/. Frequency and neighborhood density for this dataset are significantly correlated, but the correlation is extremely weak (r = 0.076, $t(2562) = 3.86$, $p < 0.05$).

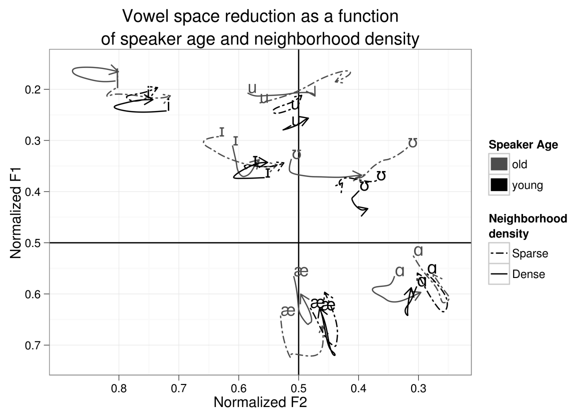

Figure 4: *Effect of neighborhood density on trajectories of vowel realizations by older and younger speakers.*

Like the midpoint analysis, word frequency affected the realization of formant contours, particularly in the back vowels. Word frequency also interacted with speaker age in the realization of /u/, with older speakers' contours extending farther back in the vowel space than those of younger speakers. Neighborhood density effects appear to be more confounded than frequency effects, either by the contextual nature of spontaneous speech or by the frequency effects themselves.

4. Discussion and conclusions

Previous research has shown that lexical factors affect the realization of vowels [1, 2]. In this study, we have extended previous approaches to a corpus of spontaneous speech. We find that lexical frequency plays a similar role in spontaneous speech as in the laboratory settings, but neighborhood density has less of a clear effect on vowel articulations. The size of both of these effects are small in terms of explained variance; however, their size is comparable to the effects on word duration in previous corpus studies, e. g. [17], and both lexical effects were larger than the variance accounted for by contextual measures, such as conditional probability. Lexical measures are more established, and predictability based on previous and following words alone may not serve capture the variation induced by context.

We find interesting interactions between the generation of the speaker and the lexical frequency of the word. Older speakers appear to be in the midst of fronting their /u/, with the more fronted version appearing in high frequency words, but younger speakers appear to have reached a more stable or completed stage of the sound change, with a smaller difference between realizations of /u/ in high and low frequency words. Other generational differences include the slight backing of /æ/ in younger speakers, though this sound change does not interact with lexical factors. These results are similar to past research on sound changes in progress [3].

Future work will expand the current study to the male participants in the corpus. Additionally, SSANOVA models will be used in future laboratory work to directly compare it to midpoint analysis in a more controlled fashion than work with corpora can accomplish. By using vowel trajectories, we may be able to see more subtle variation in vowel production than is currently possible.

5. Acknowledgements

The authors would like to acknowledge the speech production and perception seminar on for their comments and feedback.

6. References

[1] R. Wright, "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI*, J. Local, R. Ogden, and R. Temple, Eds. Cambridge University Press, 2004, pp. 26–50.

[2] B. Munson and N. P. Solomon, "The Effect of Phonological Neighborhood Density on Vowel Articulation," *Journal of Speech, Language, and Hearing Research*, vol. 47, pp. 1048–1058, 2004.

[3] C. G. Clopper and J. B. Pierrehumbert, "Effects of semantic predictability and regional dialect on vowel space reduction," *Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1682–1688, 2008.

[4] A. Bradlow, "Confluent talker-and listener-oriented forces in clear speech production," *Laboratory phonology*, vol. 7, pp. 241–273, 2002.

[5] J. M. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America*, vol. 97, pp. 3099–3111, 1995.

[6] T. M. Nearey and P. F. Assmann, "Modeling the role of inherent spectral change in vowel identification,," *Journal of the Acoustical Society of America*, vol. 80, pp. 1297–1308, 1986.

[7] R. A. Fox and E. Jacewicz, "Cross-dialectal variation in formant dynamics of American English vowels," *Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2603–2618, 2009.

[8] P. Adank, R. van Hout, and R. Smits, "An acoustic description of the vowels of Northern and Southern Standard Dutch," *Journal of the Acoustical Society of America*, vol. 116, pp. 1729–1738, 2004.

[9] C. Gu, *Smoothing spline ANOVA models*, ser. Springer Series in Statistics. Springer-Verlag, 2002.

[10] L. Davidson, "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 407–415, 2006.

[11] J. Fruehwald, "SS ANOVA," 2010.

[12] V. Chanethom, "Dynamic differences in child bilinguals' production of diphthongs." 4th ISCA Workshop ExLing 2011, 25-27 May, Paris, France, 2011.

[13] A. Baker, "Quantifying Diphthongs: A statistical technique for distinguishing formant contours." New Ways of Analyzing Variation (NWAV) 35 Conference, 9-12 November, 2006, Columbus, OH, 2006.

[14] M. A. Pitt, L. Dilley, K. Johnson, S. Kieslin, W. Raymond, E. Hume, and E. Fosler-Lussier, "Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]," 2007.

[15] B. Munson, "Lexical characteristics mediate the influence of talker sex and sex typicality on vowel-space size," in *Proceedings of the International Congress on Phonetic Sciences*, J. Trouvain and W. Barry, Eds., 2007, pp. 885–888.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]. Version 5.3.10, retrieved 12 March 2012 from http://www.praat.org/," 2012.

[17] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language*, vol. 60, pp. 92–111, 2009.

[18] M. Brysbaert and B. New, "Moving beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English," *Behavior Research Methods*, vol. 41, no. 4, pp. 977–990, 2009.

[19] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: The Neighborhood Activation Model," *Ear & Hearing*, vol. 19, no. 1, pp. 1–36, 1998.

[20] S. Gahl, "Time and thyme are not homophones: the effect of lemma frequency on word durations in spontaneous speech," *Language*, vol. 84, no. 3, pp. 474–496, 2008.

[21] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011.

[22] L. Gerstman, "Classification of self-normalized vowels," *IEEE Transactions of Audio Electroacoustics*, vol. 16, pp. 78–80, 1968.

[23] C. Gu, *gss: General Smoothing Splines*, 2012, r package version 2.0-9. [Online]. Available: http://cran.r-project.org/package=gss