INVESTIGATING VARIATION IN ENGLISH VOWEL-TO-VOWEL COARTICULATION IN A LONGITUDINAL PHONETIC CORPUS

Alan C. L. Yu*, Carissa Abrego-Collier*, Jacob Phillips*, Betsy Pillion*, Daniel Chen**

*Phonology Laboratory, University of Chicago; **ZTH Zurich
Correspondence address: aclyu@uchicago.edu

ABSTRACT

Understanding the nature of individual variation in speech, particularly the mechanism underlying such variability, is increasingly important, especially for research on sound change, since such investigations might help explain why sound change happens at all and, conversely, why sound change is so rarely actuated even though the phonetic preconditions are always present in speech. The present study contributes to the literature on inter- and intra-speaker variation in coarticulation, a major precursor to sound change, by focusing on the degree of coarticulation stressed vowels have on neighboring unstressed vowels using recordings from a longitudinal phonetic corpus of oral arguments before the Supreme Court of the United States. Significant inter-speaker variation in height coarticulation, both anticipatory and carryover, is observed, while no evidence for systematic inter-speaker variability in backness coarticulation is found. There is also no evidence for intra-speaker variation in coarticulation over the course of 205 days.

Keywords: coarticulation, corpus phonetics, inter-speaker variation, intra-speaker variation

1. INTRODUCTION

Individual variation is ubiquitous in speech. Understanding the nature of such individual variability, particularly the mechanism underlying it, is increasingly important, especially for research on sound change [16], since identifying the underlying sources of individual variability might help explain why sound change happens at all and, conversely, why sound change is so rarely actuated even though the phonetic pre-conditions are always present.

This study examines the nature of intra- and inter-speaker variability in vowel-to-vowel coarticulation in English using a longitudinal corpus of spontaneous speech. We focus in particular on the anticipatory and carryover coarticulatory effects of stressed vowels on unstressed vowels. Beddor et al. [2] show that languages have different patterns of coarticulation with respect to stress. Specifically, English stressed vowels exert strong coarticulatory effects on unstressed vowels [11, 4].

Despite these reports of stress-dependent V-to-V coarticulation, it also seems clear that this coarticulatory effect is not universal even among speakers of the same language. Magen [10], for example, found that primary stressed vowels showed a stronger effect on secondary stressed vowels only in one of four surveyed speakers. Grosvald [6] found that a great deal of inter-speaker variation in the production of V-to-V coarticulatory effects in English speakers, although even speakers whose formant differences were not statistically significant still tended to pattern in the same direction.

Beyond inter-speaker variation, the extent of intra-speaker variation with respect to coarticulatory patterns is still unknown. Previous studies on intra-speaker variation [7, 15] have largely focused on the temporal dynamics of certain speech sounds or acoustic dimensions, and paid little attention to intra-speaker variation in coarticulation per se. More recently, Zellou and Tamminga [19] examined diachronic, community-level changes in nasal coarticulation in Philadelphia English using the Philadelphia Neighborhood Corpus [8]. While they observed significant fluctuations in coarticulation throughout the four decades of the study, it remains unclear to what extent this change is observed in the same individual across time.

The present study contributes to the literature on inter- and intra-speaker variation in coarticulation by focusing on the degree of coarticulation stressed vowels have on neighboring unstressed vowels using data from the Supreme Court of the United States (SCOTUS) Corpus, a spontaneous speech corpus comprised of oral arguments before the SCOTUS.

2. THE CORPUS

The present study focuses on the recordings from the 2008-09 term of the SCOTUS Corpus. The recordings and the associated transcripts were drawn from the Oyez Project (`http://www.oyez.org/`), a multimedia archive at the Chicago-Kent College of Law devoted to the SCOTUS and its work. The 2008 term contains approximately 60 hours of audio and over 1 million words, spanning 205 days. During the hour-long oral arguments, the justices are typically very vocal participants, frequently interrupting the lawyers to ask questions, propose hypotheticals, or express disagreement. Our analysis focuses on eight Supreme Court justices; although there are officially nine justices present, one justice does not speak during this term, leaving only eight justices for analysis.

3. METHODS

3.1. Segmentation

Phone-level boundaries were determined algorithmically using the Penn Forced Aligner [18], whose acoustic models were trained on the SCOTUS corpus using the HTK toolkit [17], and which uses the CMU American English Pronouncing Dictionary [3] and includes stress (primary, secondary, or unstressed) for each vowel phone. We specifically used the FAVE (Forced Alignment and Vowel Extraction) toolkits [14].

To perform forced alignment, we used archived recordings and court transcripts in which speaker and precise speaking times are specified for each utterance. Prior to alignment, research assistants hand-checked and edited the transcripts for accuracy to the audio, and novel words were added to the pronunciation dictionary. Any interval during which multiple speakers spoke simultaneously was excluded from alignment.

3.2. Vowel measurement

Using the output from forced alignment, we used FAVE to automatically measure all vowels of duration >50 ms, taking point measurements for F1 and F2 at 1/3 duration of the vowel. An advantage of using FAVE is that the measurement algorithm is able to predict the best LPC parameter settings for each vowel (the 'Mahalanobis distance' method outlined in [5]), intended to replicate manual measurement techniques. Another key feature is remeasurement, where for each speaker in an audio file, after initially measuring all the speaker's vowels, a second pass is performed using the speaker's own means as the base of comparison for the Mahalanobis distance. With this feature, the more vowel tokens each speaker has, the more accurate the measurements. Given the large number of tokens we have per vowel per speaker, as well as the time-intensiveness of manual checking, we rely here on the automatic measurements without human correction. Formant values are normalized for speaker using the Lobanov method [9].

To examine the effects of stressed vowels on neighboring unstressed vowels, for carryover effects, we isolated vowel sequences where a target unstressed vowel (i.e. vowels labeled as AH /ə/, ER /ɚ/, IH /ɪ/, and IY /i/) is preceded by a stressed vowel (i.e. IY /i/, IH /ɪ/, AE /æ/, AA ɑ/, UH /ʊ/, UW /u/). The data set contained a total of 5620 tokens of such stressed-unstressed V-to-V sequences. For the examination of anticipatory coarticulation, we isolated vowel sequences where a target unstressed vowel (i.e., AH, ER, IH, and IY) is followed by a stressed vowel (i.e., IY, IH, AE, AA, UH, UW). The data set contained a total of 1811 tokens of such unstressed-stressed V-to-V sequences.

4. ANALYSIS

The effects of coarticulation on F1 and F2 formant values were modeled separately using linear mixed-effects regression fitted in R, using the lmer() function from the lme4 package [1]. The basic model includes BACKNESS of the stressed vowel (back (AA, UH, UW) vs. front (AE,, IH, IY)), HEIGHT of the stressed vowel (high (IH, IY, UH, UW) vs. low (AA, AE)), the quality (VOWEL: AH, ER, IH, IY) and DURATION of the target vowel. To reduce multicollinearity between predictors, continuous variables were centered, all categorical variables were sum-coded (VOWEL$_{AH}$, VOWEL$_{IY}$, VOWEL$_{IH}$, BACKNESS$_{Back}$, HEIGHT$_{High}$ = 1; as the VOWEL variable was sum-coded, the contrast with respect to ER was not tested as only three contrasts are possible.). The models also included by-subject and by-item random intercepts to allow for subject-specific and word-specific variations respectively in the specific acoustic measure. By-subject random slopes, to be specified below, were included if loglikelihood tests confirmed that the inclusion of certain types of random slopes were significant ($p < 0.05$).

We used growth curve analysis [12] to examine the intra-speaker variability of vowel production, particularly the effect of coarticulation, across days. We tested the significance of including the overall time course across the 2008 term (205 days) by modeling time with a third-order (cubic) orthogonal polynomial and fixed effects of VOWEL and contextual factors (BACKNESS and HEIGHT) on all time terms.

5. RESULTS

Figure 1 illustrates the overall effects of the backness and height of neighboring stressed vowels on unstressed vowels in our data set.

5.1. Anticipatory coarticulation

Anticipatory coarticulation on F1: The regression model for the effect of a following stressed vowel on the preceding target unstressed vowel's F1 included main effects of the DURATION and quality (VOWEL) of the target vowel and the HEIGHT of the following stressed vowel. By-subject and by-word random intercepts as well as by-subject random slopes for HEIGHT were also included. The main effect of DURATION (β=14.23, t = 7.08, $p < 0.001$) suggests that F1 is generally higher (i.e. the target vowel sounds lower) when the duration is longer. As expected, F1 value is largest when the vowel is AH, lower when the vowel is IY. A main effect of HEIGHT (β=-23.68, t = -4.65, $p < 0.001$) suggests significant vowel height coarticulation from the following stressed vowel. As illustrated in Figure 1a, F1 is lower when the following stressed vowel is high (circle) and is higher when the following vowel is low (triangle). Interestingly, the inclusion of a by-subject random slope for HEIGHT significantly increased model likelihood, suggesting the justices vary in anticipatory coarticulation in vowel height in systematic ways. None of the time terms was significant and all were excluded in the final model.

Anticipatory coarticulation on F2: The regression for the influence of a following stressed vowel on F2 included main effects of DURATION, VOWEL and BACKNESS of the following stressed vowel. Two-way interactions between DURATION and VOWEL and between DURATION and BACKNESS were also included. In addition, the model also included by-subject and by-word intercepts. By-subject random slopes for DURATION, VOWEL, and BACKNESS were not included as their inclusion did not increase model likelihood significantly.

Relative to the group mean (1804 Hz), F2 is highest when the target vowel is front, i.e., IY (β=326.3, t = 17.25, $p < 0.001$) and IH (β=229.97, t = 13.05, $p < 0.001$) and lowest when the vowel is AH (β=-236.76, t = -18.65 $p < 0.001$). While neither the main effect of VOWEL nor BACKNESS was significant, their respective interactions with DURATION were. In particular, the longer the IY vowel, the higher its F2 (β=52.92, t = 4.27, $p < 0.001$). In terms of DURATION:BACKNESS, coarticulatory backing is weaker as vowel duration increases (β=15.32, t = 2.76, $p < 0.01$). There appears to be little inter-speaker variation in the above effects as the inclusions of the corresponding by-subject random slopes did not increase model likelihood significantly. There also appears to be no significant intra-speaker variation over time, as the inclusion of the time terms did not improve model likelihood significantly.

5.2. Carryover coarticulation

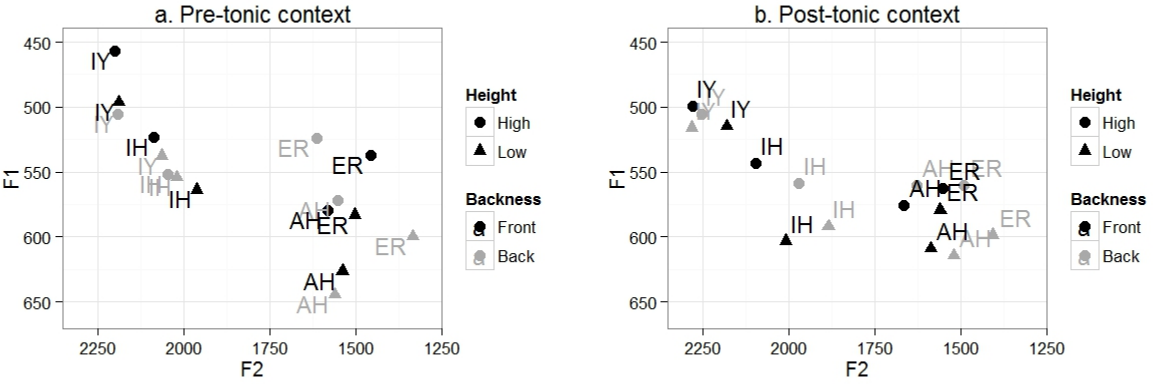
Carryover coarticulation on F1: The regression model for the effect of a preceding stressed vowel on F1 included main effects of VOWEL, DURATION and HEIGHT, as well as two interaction terms (VOWEL:DURATION and VOWEL:HEIGHT).

The mean F1 for a post-tonic target vowel is 564.63 Hz . F1 is highest for AH (β=43, t = 11.62, $p < 0.001$), and lowest for IY (β=-56.93, t = -11.17, $p < 0.001$). There is an effect of DURATION, but only as an interaction with VOWEL. The longer AH is, the higher the F1 (β=23.54, t = 5.55, $p < 0.001$); the longer IY is, the lower the F1 (β=-14.72, t = -4.22, $p < 0.001$). These results suggest a dispersion effect of duration; the longer the vowel, the more disperse a vowel is relative to the center of the vowel space. There is a significant main effect of HEIGHT, but it is vowel-specific. That is, while the preceding high vowel (the circles in Figure 1b) has a general effect of lowering F1 relative to the low vowel (triangle) context (β=-16.7, t = -6.64, $p < 0.001$), the lowering is stronger for IH (β=-10.88, t = -3.09, $p < 0.01$) and is weaker for IY (β=11.87, t = 3.13, $p < 0.01$). The HEIGHT effect on F1 is marginal for AH (β=-4.33, t = -1.75, $p = 0.08$).

The fact that including by-subject random slopes of VOWEL, DURATION, and HEIGHT and VOWEL:DURATION increase model likelihood significantly suggests that there exists important inter-speaker variation in the realization of the target vowel-specific F1 as well as the effects of context, such as vowel duration and the height of the preceding vowel, on the target unstressed vowel. No effect of intra-speaker variation is observed, since none of the time terms were significant.

Carryover coarticulation on F2: The regression model for the influence of preceding stressed vowel on F2 included main effects of VOWEL,DURATION, BACKNESS, and HEIGHT, the two-way interactions between VOWEL and the other context variables, as well as the three time terms. The model also included by-subject random slopes for DURATION and VOWEL. The inclusion of by-subject random slopes for BACKNESS or HEIGHT did not significantly improve model likelihood. The intercept, which represents the mean F2 across all post-tonic target vowels, is 1838 Hz. The effects of VOWEL are in the expected direction. AH has a lower F2 (β=-249.70, t = -18.72, $p < 0.001$), while IH and IY have higher F2 values (165 Hz and 430 Hz higher than the mean

Figure 1: F1 and F2 values of the target vowels in different stressed vowel contexts, averaged across eight justices over the entire 2008 term.



respectively). The vowel-specific F2 value changes depending on DURATION. That is, AH has a lower F2 when the vowel duration is longer (β=-34.24, t = -3.04, $p = 0.001$), while the F2 of IY is higher the longer the vowel (β29.07, t = 3.3, $p = 0.001$).

Crucially, there are significant effects of BACKNESS and HEIGHT. As illustrated in Figure 1b, in a back vowel context (gray), the F2 is significantly lower (i.e. more back) than in the front vowel context (black) overall (β=-30.15, t = -3.92, $p < 0.001$). The target vowel also has a higher F2 (β=31.05, t = 4.09, $p < 0.001$) when the stressed vowel is high (circle) and a lower F2 when the stressed vowel is low (triangle). The nature of the coarticulatory influence is also vowel-specific. The F2 increase in the high vowel context is significantly stronger in IH (β=30.18, t = 2.23, $p < 0.05$). More intriguing is the interaction between VOWEL and BACKNESS. Just in the case of the IY target, F2 is higher in the back vowel context relative to the front vowel context (β= 41.02, t = 2.71, $p < 0.01$). The fact that IY is more front in a back vowel context, on the face of it, seems like a case of dissimilation. However, further investigation reveals that many of the phonemic back vowels (i.e. UH and UW) are phonetically fronted (e.g., *jury*, *hugely*), suggesting that this might reflect a confound of regressive assimilation of frontness by the preceding stressed vowel.

Also interesting are the temporal dynamics of F2 across days. There is a significant negative linear component of time (β=-64.96, t = -2.49, $p < 0.05$), suggesting that, as a whole, F2 declines across the 2008 term. A significant positive quadratic coefficient for time suggests that F2 is lowest in mid term (β=198.23, t = 3.09, $p < 0.01$). Finally, a negative cubic coefficient (β=-132.21, t = -3.22, $p = 0.001$) suggests a high F2 toward the latter part of the term relative to the start of the term. However, the effect of time is not mediated by vowel nor by context.

The fact that the inclusion of by-subject random slopes for DURATION significantly improves model likelihood suggest that individuals vary significantly in the influence of vowel duration has on F2 realization. There are also significant difference in vowel-specific F2 across justices, as indicated by the inclusion of by-subject random slopes for VOWEL.

6. CONCLUSION

Our investigation found significant anticipatory and carryover coarticulatory effects of stressed vowels on unstressed vowels in English. There exists considerable inter-speaker variation in F1 and F2, but the nature of the variation differs depending on the coarticulatory context and the target vowel. Of particular interest is the significant inter-speaker variation in both anticipatory and carryover coarticulatory influence on F1 from the height of a neighboring stressed vowel, and yet the lack of such an effect for backness coarticulation. This state of affairs might suggest that backness coarticulation is more phonetic and mechanical, thus less prone to exhibit individual-level variation. On the other hand, vowel height coarticulation might be more planned and “phonological”, and may thus be more susceptible to different parameterization across speakers. This type of individual variation might provide the seed necessary for the eventual propagation of planned vowel-height coarticulation at the individual-level to full-fledged vowel height harmony across the speech community [13]. Finally, the finding that the coarticulatory effects do not appear to fluctuate across days at both the individual and group levels may suggest that coarticulatory effects are stable, at least across a period of 205 days.

7. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B. 2011. *lme4*. R package version 0.999375-38.

[2] Beddor, P., Harnsberger, J., Lindeman, S. 2002. Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *J. Phonetics* 30(4), 591–627.

[3] Carnegie Mellon Univ., 2008. Cmudict: The CMU Pronouncing Dictionary 0.7. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

[4] Cho, T. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *J. Phonetics* 32(2), 141–176.

[5] Evanini, K., Isard, S., Liberman, M. 2009. Automatic formant extraction for sociolinguistic analysis of large corpora. *Interspeech*.

[6] Grosvald, M. 2009. Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *J. Phonetics* 37(4), 173–188.

[7] Harrington, J., Palethorpe, S., Watson, C. I. 2000. Does the Queen speaker the Queen's English? *Nature* 408, 927–928.

[8] Labov, W., Rosenfelder, I. 2011. *Philadelphia Neighborhood Corpus*. Philadelphia: Univ. Pennsylvania, Linguistics Laboratory.

[9] Lobanov, B. V. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49, 606–608.

[10] Magen, H. S. 1997. The extent of vowel-to-vowel coarticulation in English. *J. Phonetics* 25, 187–205.

[11] Majors, T. 1998. *Stress-dependent harmony: phonetic origins and phonological analysis*. PhD thesis Univ. Texas, Austin Austin.

[12] Mirman, D. 2014. *Growth Curve Analysis and Visualization Using R*. Chapman and Hall / CRC.

[13] Paster, M. 2004. Vowel height harmony and blocking in Buchan Scots. *Phonology* 21(3), 359–407.

[14] Rosenfelder, I., Joe, F., Evanini, K., Seyfart, S., Gorman, K., Prichard, H., Yuan, J. 2014. FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.1.3. doi:10.1121/zenodo1.2935783.

[15] Sonderegger, M. 2012. *Phonetic and phonological dynamics on reality television*. PhD thesis Univ. Chicago.

[16] Stevens, M., Harrington, J. 2014. The individual and the actuation of sound change. *Loquens* 1(1), e003. doi: 10.3989/loquens.2014.003.

[17] Young, S. J. 1994. The HTK Hidden Markov Model Toolkit: Design and philosophy. Entropic Cambridge Research Laboratory, Ltd.

[18] Yuan, J., Liberman, M. 2008. Speaker identification on the SCOTUS corpus. *Proc. Acoustics* 5687–5690.

[19] Zellou, G., Tamminga, M. 2014. Nasal coarticulation changes over time in Philadelphia English. *J. Phonetics* 47, 18–35.