Interspeaker variation in anticipatory coarticulation: A whole-formant approach

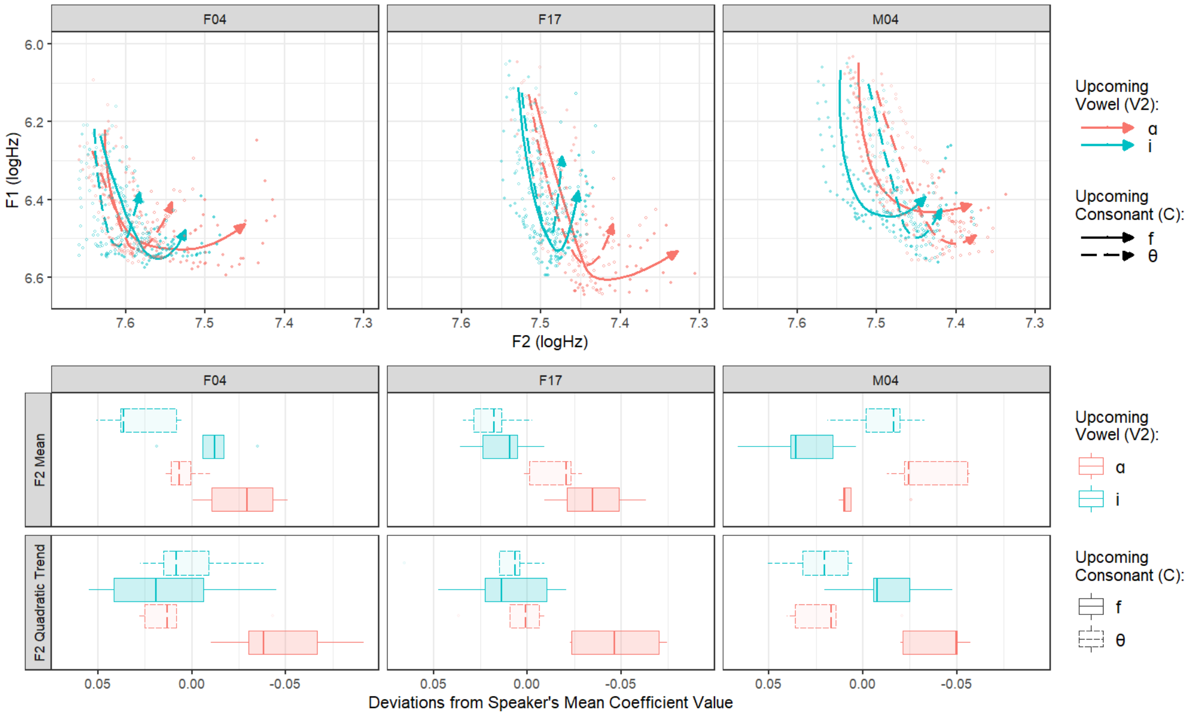
Stefon Flego, Indiana University, Bloomington (sflego@iu.edu)
Jon Forrest, University of Georgia, Athens (jrforrest@uga.edu)

Background: The body of literature studying individual differences in coarticulation identifies variability in a wide range of connected speech processes (Baker et al., 2011; Yu, 2019; Zellou, 2017; Zellou & Pycha, 2018). Most of these studies have made use of static measures of phonetic variation, but the application of time-varying representations of spectral movement to speaker-specific identification (McDougall, 2007; San Segundo & Yang, 2019; Zuo & Mok, 2015) also suggests possible variability between individuals with respect to the dynamics of connected speech processes. Both lines of inquiry suggest that expanding our measurements to include trajectory-based information can shed new light on the fine-grained differences in coarticulatory behavior between speakers. We show that whole-formant representations indeed reveal extensive interspeaker variation in the magnitude and temporal dynamics of anticipatory information in the signal.

Method: Coarticulatory data were collected from 27 speakers of American English who participated in a speech production task. Speakers produced nonce target-context word sequences containing permutations of $V_1C\#hV_2$ within a carrier sentence, in which V_1 = {ɛ}, C = {θ f}, and V_2 = {i ɛ ɑ u}, e.g. *death-heating*, *deaf-hocking*. Timing and prosody were kept consistent across speakers by entraining the syllable rate of the carrier phrase to a metronome. Nonce sequences were chosen instead of existing sequences containing the same target-context vowels to strictly control the prosodic shape of the whole carrier phrase, minimize the number of different lexical items at play, and avoid any lexical frequency effects on coarticulation that may be inherent to existing collocations. The first two formants were measured at the edges of 20 evenly spaced intervals over the course of the target vowel to approximate continuous whole-formant trajectories, to which second-degree orthogonal polynomials were fit. Each estimated orthogonal coefficient provides an independent metric for contour shape (Grabe et al., 2007; McDougall, 2007; Risdal & Kohn, 2014), so their use allows for quantitative analysis of interspeaker variability in formant dynamics.

Findings: Our data not only show a substantial amount of interspeaker variability in basic articulation of [ɛ], but also considerable interspeaker variation in the dynamics of coarticulation with upcoming gestures. To illustrate, average whole-formant trajectories for V_1 = {ɛ} preceding V_2 = {i ɑ} are shown for three of our speakers in the top row of Figure 1, along with distributions of the first and third orthogonal coefficients fit to the F2 contours of these productions in the bottom row. The distribution of the first coefficient reflects gross differences in F2, while the distribution of the third coefficient, the quadratic trend, can be thought of as reflecting different rates of acceleration in F2 over the course of V_1. In addition to differences in overall magnitude of coarticulation observed across individuals, there are robust differences in the phasing of coarticulation, with some speakers showing coarticulatory variation throughout the duration of V_1 (top right panel), while for others these coarticulatory differences only emerge relatively late in V_1's trajectory (top left panel). Furthermore, while variation in F2 attributable to the upcoming vowel is relatively similar across speakers, F2 differences due to the upcoming consonant are largely idiosyncratic (compare the way coefficients cluster with respect to upcoming consonant between M04 and the other two speakers). Finally, the relative contributions of upcoming vowel vs. upcoming consonant to spectral variation differ from speaker to speaker (e.g., the 'F2 Mean' coefficients cluster by consonant for F04 and M04, but by upcoming vowel for F17). These findings support the idea that the temporal dynamics of long-distance coarticulation vary at an individual level. The coefficients of orthogonal polynomial curves fit to whole-formant trajectories provide a means for quantitative comparison of such interspeaker differences.

Figure 1. Top Row: Average formant trajectories for [ɛ], three speakers. Bottom Row: First and third orthogonal coefficients for F2 controlling for speaker means.



References

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*(3), 347–374. https://doi.org/10.1017/S0954394511000135

Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting Intonation Labels to Mathematical Descriptions of Fundamental Frequency. *Language and Speech*, *50*(3), 281–310. https://doi.org/10.1177/00238309070500030101

McDougall, K. (2007). Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies. *International Journal of Speech Language and the Law*, *13*(1), 89–126. https://doi.org/10.1558/ijsll.v13i1.89

Risdal, M. L., & Kohn, M. E. (2014). Ethnolectal and generational differences in vowel trajectories: Evidence from African American English and the Southern Vowel System. *University of Pennsylvania Working Papers in Linguistics*, *20*(2), 138–148.

San Segundo, E., & Yang, J. (2019). Formant dynamics of Spanish vocalic sequences in related speakers: A forensic-voice-comparison investigation. *Journal of Phonetics*, *75*, 1–26. https://doi.org/10.1016/j.wocn.2019.04.001

Yu, A. C. L. (2019). On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 2. https://doi.org/10.5334/labphon.97

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13–29. https://doi.org/10.1016/j.wocn.2016.12.002

Zellou, G., & Pycha, A. (2018). The gradient influence of temporal extent of coarticulation on vowel and speaker perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 12. https://doi.org/10.5334/labphon.118

Zuo, D., & Mok, P. P. K. (2015). Formant dynamics of bilingual identical twins. *Journal of Phonetics*, *52*, 1–12. https://doi.org/10.1016/j.wocn.2015.03.003