Speech physiology or more? Vowel duration and tongue height in Northwestern Italian

Stefano Coretta, University of Edinburgh

A well-known property of vowel duration is that it tends to be modulated by vowel height: higher vowels are usually shorter than lower vowels (Lieberman 1979). A commonly accepted hypothesis regarding the source of this trend is that vowel duration is in fact related to the duration of the gesture needed to produce the vowel (assuming a consonantal starting state where the tongue body is high, Turk et al 1994): it takes less time for the tongue to go from a state of (consonantal) obstruction to the gestural configuration of high vowels (which is closer to an obstructed configuration) than to go to the gestural configuration of low vowels (which is further from an obstructed configuration). Diachronically, a coarticulatory effect of tongue height on vowel duration could be enhanced, so that tongue height alone would not be sufficient to explain vowel duration but, rather, vowel duration would be further modulated by vowel quality (Toivonen 2015, Bermúdez-Otero 2010). While the details of how this would be implemented representationally depend on the specific framework, statistically it is possible to assess the relationship of tongue height and vowel quality: if tongue height alone can explain vowel duration, the effect of vowel quality on vowel duration in a model that includes tongue height should approximate zero (this follows from the application of directed acyclic graphs for causal inference, McElreath 2019).

This study investigates vowel duration in a pre-existing data set of Northwestern Italian to assess whether the coarticulatory effect of tongue height alone is sufficient to explain vowel duration, or if an additional effect of vowel quality is necessary. Recordings from 19 speakers of Northwestern Italian (VCO province) were analysed. The recordings contain frame sentences with embedded target words of the form /CVCo/ where C is any of /p, t, k/ (the two consonants can be the same or different, in all permutations) and V is /i, e, a, ɔ, u/. Vowel onset and offset were annotated manually based on spectrographic properties following standard procedures and formant frequencies were obtained using FastTrack (Barreda 2021). F1 is used as a proxy for tongue height. A Bayesian regression model was fitted to logged vowel duration, with the following predictors: a parametric effect of vowel (to model differences in average vowel duration by vowel), a smooth term over standardised F1 (to model the potentially non-linear effect of F1) and a factor smooth interaction over F1 by speaker with vowel as a by-variable (to model by-speaker differences).

The results indicate that an effect of F1 coexists with an additional effect of vowel quality, as illustrated in Figure 1 (the predictions for each vowel have their own intercept, indicating that F1 is not sufficient to explain vowel duration). Diachronically, this means that the putative original physiological effect has been enhanced so that the total effect is not that which is expected by physiology alone. This could be interpreted as a case of entrenchment and a special type of rule scattering (Bermúdez-Otero 2010), where a physiological "rule" coexists with a "phonetic" one.

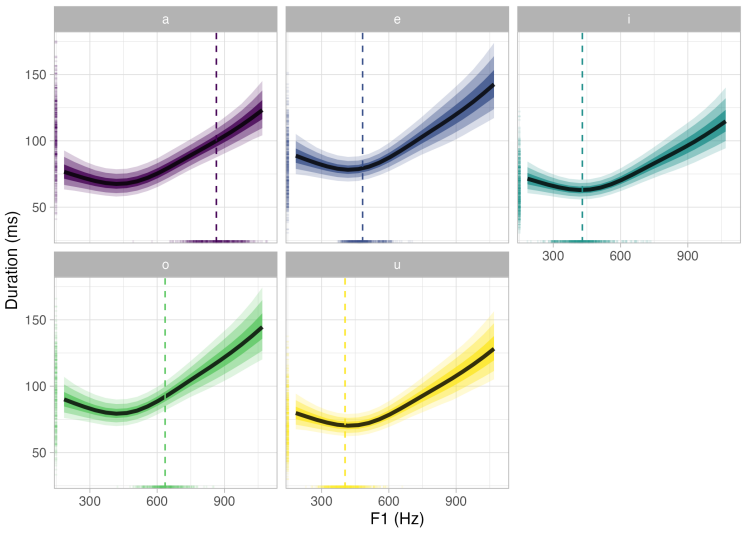

Figure 1. Predicted vowel duration (y-axis) as a function of F1 (x-axis) and vowel quality. The vertical dashed lines indicate the mean F1 value of each vowel as found in the data set. The short coloured lines on the axes are the F1 and duration values in the raw data.