

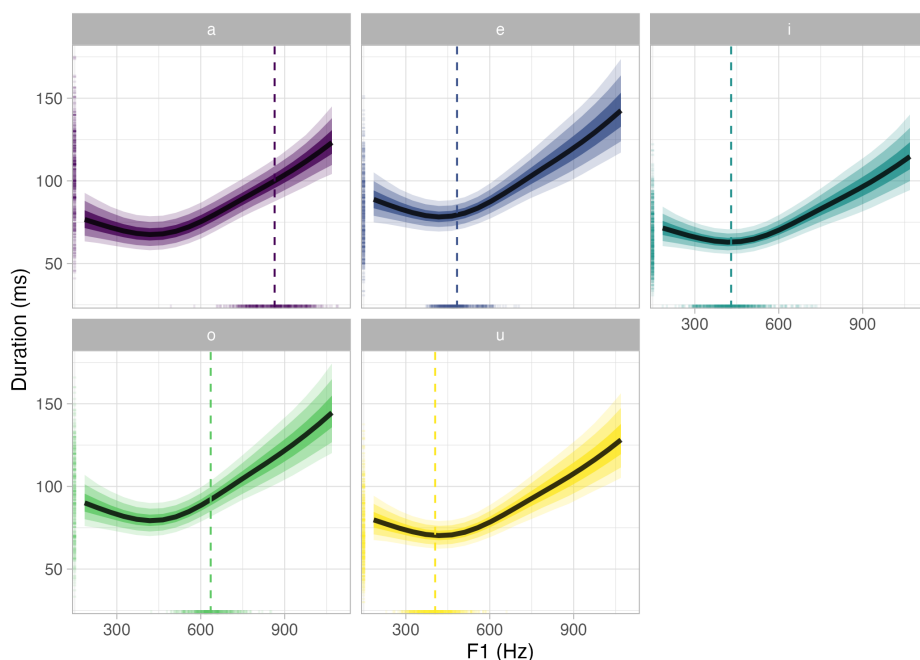
# Is vowel duration the product of gestural timing alone? Data from Northwestern Italian

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 higher vowels (which is closer to an obstructed configuration) than to go to the gestural configuration of lower vowels (which is further from an obstructed configuration). Diachronically, a coarticulatory effect of (gradient) 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 depending on the vowel category (Toivonen 2015, Bermúdez-Otero 2010). While the details of how this would be implemented synchronically (representationally) depend on the specific framework, statistically it is possible to assess the nature of the relationship between vowel duration, tongue height and vowel: **if (gradient) tongue height alone can explain vowel duration, the effect of vowel category 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 (Verbano-Cusio-Ossola 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 term for vowel category (/i, e, a, ɔ, u/, sum-coded; to model differences in average vowel duration by vowel), a smooth (non-linear) 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 parametric term of vowel category is the critical term: **if the coefficients of this term are robustly different from 0, then we can say there is evidence for a vowel quality effect in addition to the F1 effect.**

The 95% Credible Intervals of the vowel category coefficients (/a/ [-0.38, +0.02], /e/ [+0.21, +0.37], /i/ [-0.48, -0.32], /ɔ/ [+0.26, +0.41]) strongly suggest that the effect of vowel quality is non-null, even when accounting for the effect of F1. This means that F1 by itself is not sufficient to explain vowel duration. **The results thus indicate that an effect of F1 on vowel duration coexists with an additional effect of vowel quality.** Figure 1 shows the model predictions of vowel duration for each vowel, depending on F1. F1 has a non-linear effect on vowel duration, by which the effect is negative up to about 500 Hz and positive from around 600 Hz. In addition, the relationship between F1 and vowel duration is further modulated depending on the vowel: this can be observed in the plot as the different heights of the regression lines. **In conclusion, the data at hand provide evidence for an effect of vowel quality on vowel duration in addition to an effect of F1 in Northwestern Italian.**



**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 ticks on the axes are the F1 and duration values in the raw data.