Generalised additive mixed modelling for dynamic formant analysis

Márton Sóskuthy*, University of York*

A common strategy in acoustic studies of vocalic sounds is to take formant measurements at a single point for each target sound (e.g. Peterson & Barney, 1952; Labov, Ash & Boberg, 2005; Hay, Pierrehumbert, Walker & LaShell, 2015). This method inevitably results in a loss of information about time-varying aspects of speech production. However, it is undeniable that it has a number of practical advantages: it allows the analyst to use common and well-documented tools for data analysis (e.g. linear mixed effects modelling), it minimizes the influence of surrounding sounds and it makes it easy to create familiar scatterplots of vowel productions. Single-point measurements are also relatively easy to implement as part of systems that carry out automated vowel analysis with the help of forced-alignment (Fromont & Hay, 2012; Rosenfelder et al., 2014), and they tend to be relatively accurate even without manual correction by human analysts (Evanini, 2009).

On the other hand, formant analysis relying on multiple measurements per target sound (i.e. dynamic formant analysis) may yield a more accurate picture of time-varying properties of speech sounds, but it presents a number of analytical challenges. Dynamic formant analysis is a relatively new but increasingly popular approach (e.g. Watson & Harrington, 1999; Fox & Jacewicz, 2009; Cardoso, 2015). As a result, there are no widely accepted methodological standards, and analyses by different teams often share little common ground. It is not always clear in what way a dynamic analysis is superior to a single-point analysis, and what method one should use to answer a specific question related to speech dynamics.

In this talk, I present a detailed overview of one specific method for dynamic formant analysis: generalised additive mixed modelling (Wood, 2006). The goal of this talk is not to argue that additive models provide a one-size-fits-all solution for dynamic formant analysis. Instead, I will attempt to clarify what types of questions additive models allow us to answer, and bring attention to a number of methodological issues that need to be considered when using them. Two problems that will be discussed in detail relate to appropriate random effects structures for additive models fit to formant trajectories and the use of confidence intervals in evaluating the results of additive models. I will illustrate these points using Monte Carlo simulations and data from projects looking at (i) /u/-fronting in Derby English (Sóskuthy et al., 2015) and (ii) coda liquids in Glasgow English (Stuart-Smith et al., 2015).

References

Cardoso, A. (2015). Variation in nasal–obstruent clusters and its influence on PRICE and MOUTH in Scouse. *English Language and Linguistics*, *19*(03), 505-532.

Evanini, K. (2009). *The permeability of dialect boundaries: a case study of the region surrounding Erie, Pennsylvania*. PhD Dissertation, University of Pennsylvania.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America*, *126*(5), 2603–2618.

Fromont, R., Hay, J. (2012). LaBB-CAT: An annotation store. In *Proceedings of Australasian Language Technology Association Workshop* 10. 113–117.

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91.

Labov, W., Ash, S., & Boberg, C. (2005). *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H. & Yuan, J. (2014). FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2. DOI: 10.5281/zenodo.22281.

Sóskuthy, M., Foulkes, P., Hughes, V., Hay, J., Haddican, B. (2015). Word-level distributions and structural factors codetermine GOOSE fronting. *Proceedings of the 18th International Congress of Phonetic Sciences*, 10-14 August 2015, Glasgow.

Stuart-Smith, J., Lennon, R., Macdonald, R., Robertson, D., Sóskuthy, M., José, B., Evers, L. (2015). A dynamic acoustic view of real-time change in word-final liquids in spontaneous Glaswegian. *Proceedings of the 18th International Congress of Phonetic Sciences*, 10-14 August 2015, Glasgow

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458–468.

Wood, S. (2006). *Generalized additive models: an introduction with R*. Boca Raton: CRC press.