# Statistics of vowel formants: Normality of distributions across the syllable

D. H. Whalen
City University of New York; Haskins Laboratories; Yale University

Variability has been a topic in phonetics from the first time we were able to quantify the acoustic signal. Our ability to analyze ever larger datasets has opened new possibilities for refinement of our understanding. One area where this can be applied is the study of vowel formants, as a way of understanding the targets and online control of vowel articulation. One immediate question is: Are vowel formants normally distributed? Most acoustic studies cannot begin to answer this question, due to the small number of repetitions usually collected. Twenty seems like a substantial number, and for the purposes of finding a central tendency, twenty is a fairly good number. However, to have any idea whether that distribution is normal or not requires a much larger sample. One such case will be studied here.

A caveat about formant measurements is in order before laying out the experiment. We have known for some time that our formant measurements are biased toward the nearest harmonic (Klatt, 1986; Vallabha & Tuller, 2002). (We are really interested in the resonance, not the formant; Titze et al., 2015.) Shadle, Nam and Whalen (2016) demonstrated that the errors are still present; the result I will present used the LPC algorithm of Praat (Boersma, 2001), which is sensitive to that artifact, but no other system, including the more accurate Weighted Linear Prediction with Attenuated Main Excitation (WLP-AME; Alku, Pohjalainen, Vainio, Laukkanen, & Story, 2013) was automated enough for use with a large dataset.

In order to have enough tokens to establish the normality of the vowel formant distribution, it was necessary to have a speaker produce many repetitions of the same word without saturating the production system. The 1,000 repetitions of the word "bucket" in Kello et al. (2008) virtually guaranteed more variability than we would expect from a more distributed set of utterances. Here, four target English forms were used ("heed," "geek," "ode/owed," and "dote"). These consisted of environments that were expected to have small amounts of acoustic results of coarticulation ("heed" and the homonyms "ode"/"owed") and large amounts ("geek" and "dote"). (The homonyms "ode" and "owed" were included for ancillary reasons; each of them occurred half as often as "heed.") These items were randomized together with 200 filler words. 100 repetitions of the target words were collected in each recording session. The filler words occurred once in the first half of the session, and once in the second half. Within each sequence of 8 items, four target words and four filler words appeared in random order. The homonyms alternated so that each occurred 50 times per session. So far, 200 of a planned 500 repetitions have been collected.

Initial analyses via a Gaussian mixture model (Berge, Bouveyron, & Girard, 2012) indicate that formant values at 40% of the duration of the vocalic segment come from a single (normal) distribution. This was true for both the relatively uncoarticulated and the highly coarticulated environments, as can be seen in the stable standard errors of the SSANOVA of the first two formants (see Fig. 1). These results will be compared with those of Niziolek et al. (2013), who found within-syllable corrections across 200 repetitions of English words. The present results instead support the dynamic vowel and its resultant formant trajectory as the unit of planning.

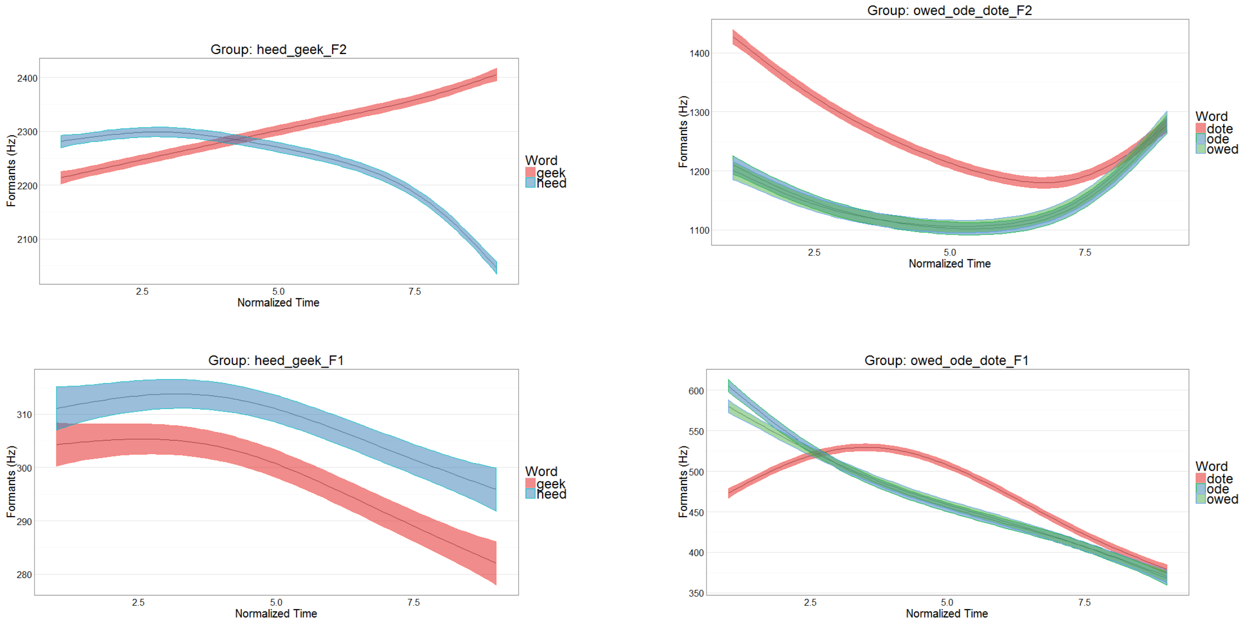

Figure 1: SSANOVA plots of F2 (top) and F1 (bottom) for 200 repetitions of “heed” and “geek” (left) and “ode”/”owed” and “dote” (right).

Alku, P., Pohjalainen, J., Vainio, M., Laukkanen, A.-M., & Story, B. H. (2013). Formant frequency estimation of high-pitched vowels using weighted linear prediction. *Journal of the Acoustical Society of America, 134*, 1295-1313.

Berge, L., Bouveyron, C., & Girard, S. (2012). HDclassif: An R package for model-based clustering and discriminant analysis of high-dimensional data. *Journal of Statistical Software, 46*(6), 1–29.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341-345.

Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cognitive Science, 32*, 1217-1231. doi:10.1080/03640210801944898

Klatt, D. H. (1986). Representation of the first formant in speech recognition and in models of the auditory periphery. In P. Mermelstein (Ed.), *Proceedings of the Montreal satellite symposium on speech recognition, 12th International Congress on Acoustics* (pp. 5-7). Montreal: Canadian Acoustical Society.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience, 33*, 16110-16116.

Shadle, C. H., Nam, H., & Whalen, D. H. (2016). Comparing measurement errors for formants in synthetic and natural vowels. *Journal of the Acoustical Society of America, 139*, 713-727.

Titze, I. R., Baken, R. J., Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., . . . Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *The Journal of the Acoustical Society of America, 137*(5), 3005-3007. doi:doi:http://dx.doi.org/10.1121/1.4919349

Vallabha, G. K., & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication, 38*, 141-160. doi:doi: 10.1016/S0167-6393(01)00049-8