

**Reconsidering Vowels as Mathematical and Statistical Entities:
How Much Variation Should There Be?**

The presentation and discussion of vowel data in most modern dialect descriptions is based on format values averaged across tokens, environments, and speakers. Though this method is known to obscure much interesting intra- and inter-speaker variation, dialectologists and sociolinguists accept this loss of information for the trade-off of a more easily interpreted data set. Because vowels are treated as mathematical and statistical entities, this averaging of values is considered an acceptable way to apply and disseminate dialect findings.

However, when the data points for vowels are not combined, when the means for individual tokens, speakers, and communities are considered in full, a more diverse picture emerges. For example, among speakers in Southern Illinois (a South Midland dialect region), the amount of variation both within and between speakers for any given vowel category is striking.

Consider, for example, Figure 1, below, which shows male speakers' productions of the GOOSE or /ʊ/ vowel along the F2 dimension. In this figure, the mean F2 of GOOSE from word list data only is plotted for each individual male high school speaker; the backest variants are indicated with large Xs, the frontest variants are indicated with a large open squares (in this figure, the Y-axis represents nothing; it is non-dimensional).

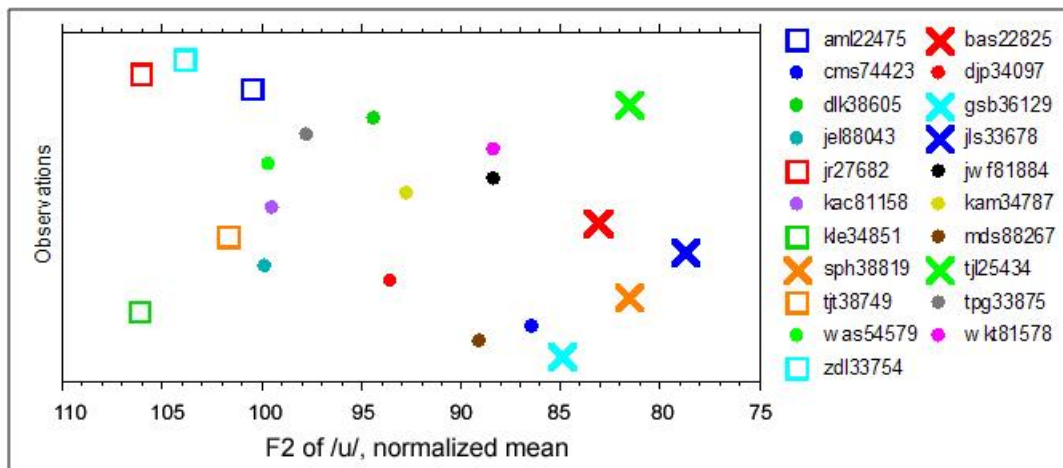


Figure 1: Mean Productions of the F2 of /ʊ/ for High School Males.

Furthermore, even this is not the whole picture—the data points in Figure 1 are themselves averages across ~10 tokens of the GOOSE vowel for each speaker. With this in mind, our question becomes not only how we might explain this level of variation via sociolinguistic means, but also whether or not treating the perceptual and indexical entities that vowels are as mathematical units remains a safe assumption.