Evidence for vowel targets in formant distributions and within-syllable adjustments

D. H. Whalen
Program in Speech-Language-Hearing Sciences/Dept. of Linguistics, City University of New York; Haskins Laboratories; Dept. of Linguistics, Yale University
whalen@haskins.yale.edu

Variability is an inescapable aspect of human language yet languages continue to serve their function of conveying content and indexical information. While some variability is governed by context (such as the coarticulatory influences of neighboring phones), some variability is presumed to be random. Whether this latter randomness is due to noise in neural systems, intrinsic limitations of the speech production mechanism, or other factors has not been determined. Several proposals have been made attempting to account for systematic variation, including coarticulation limited by time (Bell-Berti & Harris, 1981), size of vowel inventory (Manuel, 1987) or regions of influence (Keating, 1990). The random variability is more difficult to study, but Tilsen (2015) showed that variability occurs even when every effort, including monetary reward, is made to reduce it. Just how this variability aligns with the target is an open question. If it is normally distributed, it indicates a central target. If the distribution becomes flat between two values (all values in a range are equiprobable), it would indicate that the target is, instead, a region.

Recent neuroimaging and acoustic evidence has led to a proposal that ideal targets exist even when the execution is not ideal (Niziolek, Nagarajan, & Houde, 2013). Though vowel productions outside the typical range for the intended vowel elicited the kinds of neural responses that are associated with overt slips of the tongue, their formants were found to tend toward the typical area, even though the time scale (50 ms) was too little to allow for auditory feedback to play a role, suggesting an influence of a target on variability.

The present experiment explores the relationship between vowel targets and the resultant variability by measuring a large number of productions that are separated by as much time as possible. 500 repetitions of four target CVCs (5 words) (“heed”, “owed/ode”, “geek”, “dote”) are collected across five days. Each day, a different pseudo-randomized list is read, with 100 of each CVC (thus 50 each of “owed” and “ode”) interspersed among an equal number of filler items. Filler items are 200 (C)VVC English words; each of these is presented twice in each randomization. Every 8 words includes one exemplar of the target CVC, along with a random selection of 4 filler items. Filler items do not repeat until the second half of the experiment. Thus even though the words are said many times, they are never said in succession. The list of words is practiced before recording to ensure that all words are familiar to the speaker. Stimuli are displayed on a computer screen via the Presentation program (www.neurobs.com). The speaker is instructed that this is not a reaction time experiment, so that production should begin only when the speaker is sure that the word is going to be produced correctly. F0, intensity and duration are to be kept as constant as possible. Four blocks of 200 items are presented, with a break between blocks. Extra, non-analyzed words begin every block. The speaker is a practiced radio announcer and phonetician.

Formant analysis for this preliminary assessment was performed with the Burg analysis in Praat (Boersma & Weenink, 2009), with settings of a time step of 10 ms, 5 formants with a maximum value of 5000, window length of 2.5 ms, and preemphasis
beginning at 50 Hz. Future analyses will be conducted with alternative measures that are less sensitive to the location of the harmonics of the F0 (Klatt, 1986).

Preliminary analysis of the first session (i.e., 100 repetitions per CVC) indicates that the distribution is likely to be normal, but this is too few values to draw firm conclusions. Changes within a syllable only partially replicated Niziolek et al. (2013). Their results (Fig. 1) showed consistent shifts toward the means. The present results (Fig. 2) were more varied. While 71% of F1 shifts and 65% of F2 were in the predicted direction, 58% of the F1 and 35% of the F2 trajectories overshot the mean (lines cross the black lines in Fig 2.)

The results should shed light on the underlying target for vowels and the frequency with which online adjustments are made toward that target.

![Fig. 1](image1.png) ![Fig. 2](image2.png)

Fig. 1: Formant shifts in Nizolek et al. (2013:16114). If the formants started in the periphery (left, red), they tended toward the center; otherwise (green), they tended to remain the same.

Fig. 2: Results for 100 tokens of “heed.” F1/F2 trajectories marked by arrows (red, F1 < mean F1; blue, F1 > mean F1); arrowhead is at midpoint of the vowel. Mean F1 and F2 shown with black line.

References: