How informative is dialect about vowel distributions?

Dave F. Kleinschmidt, Kodi Weatherholtz, T. Florian Jaeger

Listeners need to cope with variability across talkers for successful speech perception. This can be framed as a problem of statistical inference: in order to infer a talker's intended category from observed cues, the listener needs to know the distribution of cues for each category. But these distributions vary across talkers. When viewed this way (as in the ideal adapter framework, Kleinschmidt & Jaeger, 2015, *Psych.Rev.*) talker variability requires inference at another level: listeners must *infer* the current talker's cue distributions, by combining the statistics of the current speech with prior experience with other talkers. Critically, this implies that listeners should use any available information that's informative about an unfamiliar talker's cue distributions. The *structure* of the variability across talkers will determine which factors are informative. For instance, if talkers primarily vary based on socio-indexical groups like gender or regional dialect, listeners can draw on experience with other talkers from the same group when adapting. However, if talkers vary in purely idiosyncratic ways, listener should adapt to each talker individually, not drawing on specific experience with any other particular talker or group.

Sociolinguistic research suggests that grouping factors such as gender and dialect are indeed informative about phonetic variation. What is lacking, however, is a quantitative framework that allows us to compare the *relative* informativity of different grouping factors for any given pattern of phonetic variation. Here we begin to develop and evaluate such a framework, based on the ideal adapter. Specifically, we consider four ways of grouping observed category tokens: by individual talker, by gender, by dialect, and by both gender and dialect. We assess the consequences of attending to versus ignoring these different socio-indexical grouping factors during speech perception. We calculate these consequences at two levels:

1. **Informativity** at the level of cue distributions: the difference between the marginal (or average) distribution across talkers and the corresponding distribution conditioned on each grouping factor.
2. **Utility** at the level of category recognition: the probability of correct recognition of individual tokens by classifiers with knowledge of different distributional statistics (e.g., dialect- or gender-specific statistics).

Data. We base our analysis on the first and second formant values (F_1 and F_2) measured from isolated productions of hVd words in the Nationwide Speech Project corpus (Clopper, Pisoni, & de Jong, 2005, *JASA*). This corpus contains productions from 48 talkers: 8 talkers (4 male and 4 female) from each of 6 major dialect regions of the United States (New England, Mid-Atlantic, North, Midland, South, and West). We assess whether considering dialect-level groupings of talkers is informative, and specifically whether dialect-specific distributions are informative at an *intermediate* level between language-wide (marginal) and talker-specific distributions.

Informativity at the level of cue distributions. We used the Kullback-Leibler (KL) divergence to assess the relative informativity of grouping factors with respect to cue distributions. KL divergence is an information-theoretic measure of the information lost when using one distribution (e.g., the marginal or average distribution of acoustic cue values) to estimate another distribution (e.g., the corresponding talker- or group-specific distribution). Figure 1 shows the KL divergence (information loss in bits) when using the marginal $F_1 \mathrm{x} F_2$ distribution across all talkers in the NSP corpus rather than distributions conditioned on (i) dialect, (ii) gender, (iii) both dialect and gender, and (iv) talker.

The information loss is significantly larger than zero in all cases (as indicated by the bootstrapped confidence intervals). This indicates that each of the grouping factors carries information about the distribution of $F_1 \mathrm{x} F_2$ values. However, the relative *amount* of information loss varies considerably across grouping factors. Talker identity carries the most information, as indicated by the relatively large KL divergence, but dialect also carries information, even above and beyond gender.

These results are in line with sociolinguistic research: dialect and gender affect the realization of vowels. Going beyond this research, we can now quantify the relative influence of these socio-indexical factors with respect to perception and category recognition.
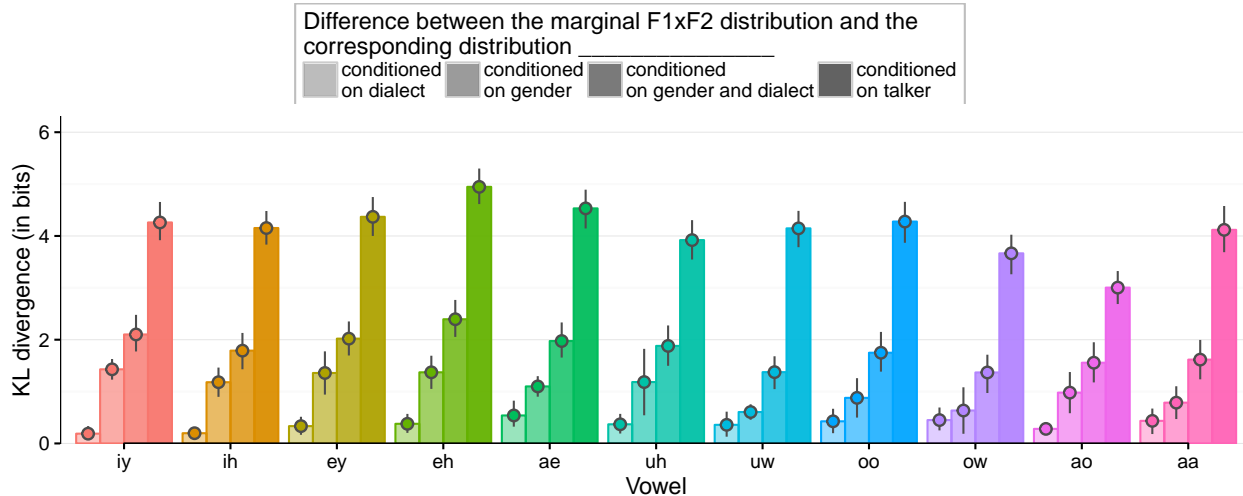
Figure 1: KL divergence between marginal formant distributions and each grouping level's distribution.

Utility at the level of category recognition. Different cue distributions lead to different optimal phonetic categorization strategies. It is possible, however, that this variation does not cause categorization *errors* often enough to matter in perception. The *utility* of any socio-indexical grouping factor thus depends on the extent to which categorization errors are reduced by drawing on knowledge of how variation is structured with respect to that group. As a measure of utility, we assessed the probability of correctly recognizing each vowel token under the marginal $F_1 x F_2$ distributional statistics or the corresponding statistics conditioned on the socio-indexical factors of interest. Figure 2 shows the results. As with the informativity, the utility is highest for the talker-specific distributions. The importance of dialect varies across vowels: for some its utility is nearly as high as gender, while for others it contributes little if anything.
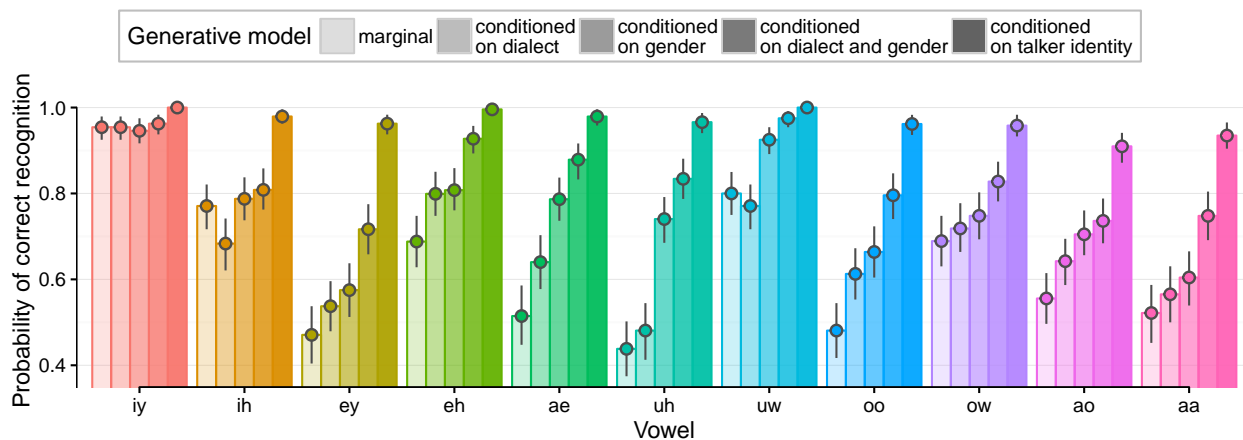

Figure 2: Comprehension accuracy for each vowel using different levels of grouping

Conclusions. We have assessed *how* variability in vowels is structured. Specifically, we assessed the relative amount of information that different socio-indexical grouping factors carry with respect to vowel variation. Moreover, using an ideal observer model, we have shown that knowledge of how vowel variation is structured substantially improves comprehension in the face of talker variability. With this model in mind, we are equipped to predict which factors listeners should be sensitive to when adapting, making (for the first time) quantitative predictions about the relative importance of different socio-indexical factors for perception.