A Bayesian belief updating model of phonetic recalibration and selective adaptation

Dave Kleinschmidt¹ and T. Florian Jaeger²

University of Rochester Brain and Cognitive Science¹,² and Computer Science²

contact: {dkleinschmidt,fjaeger}@bcs.rochester.edu

The mapping from phonetic categories to acoustic cue values is highly flexible, and adapts rapidly [6] and robustly [3] to exposure. To our knowledge, there is not currently any theoretical or computational framework which explains the range of these effects. We have developed a novel Bayesian belief updating model which naturally unifies two of the most commonly studied types of adaptation (selective adaptation and phonetic recalibration) which are usually considered to be distinct.

Vroomen et al. [6] used audio-visual speech stimuli to induce changes in subjects’ b-d categorization behavior. Acoustic tokens from a /b/-to-/d/ continuum were paired with videos of a speaker articulating /b/ or /d/. When repeatedly exposed to acoustically ambiguous /b/ stimuli paired with unambiguous /b/ videos, subjects categorized more items on an acoustic continuum as /b/, and likewise classified more items as /d/ after ambiguous-/d/ exposure. However, after exposure to acoustically unambiguous /b/ audio-visual stimuli, subjects classified fewer items as /b/, and likewise for /d/.

The first result is an example of phonetic recalibration, where a phonetic category “expands” to encompass previously ambiguous tokens, and the second is an example of selective adaptation, where repeated exposure to unambiguous tokens from a category causes the exposure category to “shrink”. These two types of adaptation, which have both been studied extensively, have been generally considered to be distinct, with selective adaptation analyzed as a sort of receptor fatigue or other transient process [1] and recalibration as a restructuring of underlying representations [3]. Indeed, Vroomen et al. [6] found that they build up over different time courses (Figure 1, dashed lines), with selective adaptation (bottom) increasing in strength with further exposure while recalibration (top) occurs quickly but eventually fades, suggesting that different underlying mechanisms are at work.

However, we have developed a unified model which predicts both types of adaptation. In this model, a listener’s phonetic categories are modeled as normal distributions over phonetic cues, and their beliefs about those categories as distributions over the category means and variances. After exposure to items from a particular category, the listener’s beliefs about that category are updated (via Bayesian inference) to bring their representation into better alignment with their recent experience.

This model is thus a computational model of the cognitive processes underlying the flexible mapping from phonetic categories to corresponding cues. While this model is not itself a process model, such models exist which approximate Bayesian belief update under cognitively-realistic constraints (e.g. [5]). Based on general assumptions, our model straightforwardly predicts the aggregate pattern observed by Vroomen et al., including the specific time course of selective adaptation and recalibration (Figure 1), as well as the behavior of individual subjects (Figure 2). Beyond simply providing a good fit to subjects’ behavior, this model also suggests that phonetic adaptation happens at a multimodal level or representation, since a version of the model which treats phonetic categories as distributions over purely acoustic cues fails to capture the time course of recalibration.

Furthermore, through hierarchical extensions, this model has the potential to account for patterns of generalization of phonetic adaptation between speakers and categories (e.g. [4, 2]). This model thus constitutes a first step towards a novel and general theory of phonetic categories which respects the flexibility of these categories without discarding structured representations altogether.
Figure 1: Overall fit of belief updating model \( (R^2=0.67) \) to data of Vroomen et al. [6], showing the time course of recalibration to ambiguous stimuli (red) and selective adaptation to unambiguous stimuli (blue). Solid lines correspond to the best fit averaged over subjects, and dashed lines correspond to empirical difference scores (percent /b/ responses after /b/ exposure minus percent /b/ responses after /d/ exposure), with shaded 95% CIs of the means.

Figure 2: Best model fit for each individual subject. Dashed lines are empirical difference scores (shaded regions are 95% confidence intervals) and solid lines are the best-fitting model for that subject. Mean \( R^2=0.57 \), SE=0.04.

References:


