Mixture models of phonologization

Introduction. PHONOLOGIZATION (Hyman, 1976) is often invoked to explain the transition from automatic to controlled speech gestures, such as the emerging importance of $f0$ as a cue to initial stop identity in Seoul Korean (Kang & Guion, 2008). However, the model is currently *post hoc* rather than predictive. In particular, it cannot explain why certain gestures are targeted for enhancement and not others, or under what conditions this enhancement is likely to occur.

Proposal. The factors influencing the selection and enhancement of cues in phonologization may be understood in a MIXTURE OF GAUSSIANS framework, in which subphonemic variation is encoded directly in the phonological representation (Pierrehumbert, 2001; Solé, 2003). This representation suggests that phonologization may be an adaptive response to systematic bias: cues are enhanced proportional to their informativeness in distinguishing a contrast and the current precision of that contrast, a strategy which optimizes both speaker- and listener-oriented constraints (Lindblom et al., 1995). The likelihood of phonologization vs. merger may be predicted based on the overall log-likelihood of the model for some data. We implement this proposal and show its application to the case of $f0$ phonologization in Seoul Korean.

Model and simulations. Agent-based simulations were conducted in which $D = 5$ cue dimensions relevant for the perception of Korean lenis and aspirated stops (VOT, vowel length, closure duration, $H_1 - H_2$, and $f0$) were represented using a Gaussian mixture model with $K = 2$ components. At each timestep, agents produced an exemplar $\mathbf{x_k} = \{x_1, \ldots, x_D\}$ of component k by sampling from a 5-dimensional density $\mathcal{N}\pi_k(k; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, using maximum likelihood parameter estimates based on previously experienced tokens. The probability of enhancing cue d at each step was a function of (i) a cue's WEIGHT $\omega_d \in [0, 1]$, given as the difference between category means over their variance in that dimension, and (ii) the contrast PRECISION, defined as the error rate of a maximum *a posteriori* classifier (Nearey & Hogan, 1986). The degree of enhancement was similarly proportional to these parameters. Tokens were also potentially subject to a BIAS λ on d, after which they were assigned a label with a probability equal to the relative likelihood of each component, calculated as the combination of conditional probabilities $p(k|x_1), \ldots, p(k|x_D)$. At the end of each timestep, the number of components K was increased or reduced to reflect the model with the lowest overall log-likelihood.

Results. Simulations of up to 50,000 iterations were conducted with and without both enhancement and systematic bias, using Korean data from the apparent time study of Kang & Guion (2008) as the initial and target distributions. Neither the proposed enhancement strategy nor systematic bias alone were sufficient to induce phonologization of $f0$, but application of both gave a close approximation of the attested distributions and cue weights, as measured by the Kullback-Leibler divergence between the simulated results and the empirical targets. In a second series of simulations where the weights of secondary cues to the lenis/aspirated contrast were equal at initialization, neutralizing bias in the production of VOT (the primary cue) led to category merger or stability of the existing cue structure, depending on bias strength.

Conclusions. Phonologization may result from the interaction of probabilistic enhancement and systematic bias in the production of a primary cue, depending on bias strength and the initial state of subphonemic cue weights. Potential cue targets in phonologization may be predicted based on their relative weight in a Gaussian mixture model, which together with contrast precision, can predicte the global likelihood of phonologization. In concert with an enhancement strategy, bias can drive phonologization of an existing covert contrast, but if all secondary cues are of similar weight, contrasts are predicted to either collapse or remain stable.