**Learning the context-dependent perception of novel speech sounds**

Masaki Noguchi and Carla Hudson Kam
University of British Columbia

When a sound is produced in different contexts, the acoustic signal associated with that sound can show a significant amount of variation due to coarticulation with the surrounding sounds. Despite this, listeners are able to establish a single percept by taking information from the contexts into consideration and potentially factoring out the aspects of the variation that can be attributed to the coarticulation [1]. We are interested in how listeners come to have this ability, that is, to perceive speech sounds by integrating acoustic information from the sounds and their contexts.

Recent studies have suggested that contextual cue integration plays an important role in the learning of phonological status [2, 3]. In our previous study, we exposed native English-speaking adults to input in which the tokens of two novel sounds, retroflex [ʂa] and alveolopalatal [ɕa], showed a frequency profile known to lead to the learning of phoneme-like categories [4], but occurred in mutually exclusive contexts. After exposure, the participants showed reduced sensitivity to the contrast between the novel sounds, suggesting that they learned the novel sounds as allophone-like categories. This change in sensitivity, however, only occurred when the pattern of the complementary distribution was phonetically "natural", that is, when there were phonetic similarities between the sounds and their respective contexts; i.e. retroflex [ʂa] occurred after [u] and alveolopalatal [ɕa] occurred after [i] [3].

A possible explanation for this asymmetry is that the participants lost their sensitivity to the contrast between the sounds not only because the sounds were presented in complementary distribution, but also because the connections between the sounds and their respective contexts induced contextual cue integration in such a way that the perceptual distance between the sounds became smaller than the acoustic distance between the sounds. The contrast between retroflex [ʂa] and alveolopalatal [ɕa] is cued by F2 transition, [ʂa] having a lower F2 onset than [ɕa] . When a token of [ʂa] is presented after [u], however, the low F2 onset of the token can be analyzed as a result of coarticulation with the preceding [u] and the token may sound less retroflex. Similarly, when a token of [ɕa] is presented after [i], the high F2 onset of the token can be analyzed as a result of coarticulation with the preceding [i], and the token may sound less alveolopalatal. In this way, perceptual distance between the sounds can be reduced if presented in natural contexts, which would interfere with learning or maintaining the distinction between the sounds.

In this study, we investigated the possibility that the learners in our previous study might have learned the contextual cue integration as they learned two novel sounds. We assessed the perception of the same novel sounds in the same natural and unnatural contexts by native English-speaking adults before and after exposure to the same learning stimuli as in our previous study. If the contextual cue integration was learned via exposure, participants should perceive the sounds as being more similar to each other in natural contexts than in unnatural contexts, but only after exposure to the learning stimuli. In contrast, if the contextual cue integration is inherent to auditory processing, participants' perception of the sounds should be dependent on the contexts even before the exposure.

**Method:** 20 adult native English speakers participated in the study. The experiment consisted of two sessions over two consecutive days. In session 1, participants performed a similarity rating task first, then listened to ~15 mins of input. In session 2, participants listened to the input first, then did another similarity rating task. The input comprised 512 bisyllabic strings. Half of the strings contained tokens of novel sounds, and the rest were fillers. Novel sound tokens were 8 distinct syllables taken from a 10-step continuum between [ʂa] and [ɕa]; 4 syllables from each side of the category boundary were selected. The frequencies of these syllables were manipulated so that their aggregate distribution showed a bimodal shape with a frequency peak on each side of the category boundary. The novel sound tokens were presented in both natural and unnatural contexts: both after a syllable with vowel [u] and a syllable with vowel [i]. In the similarity rating task, participants rated the similarity of [ʂa] and [ɕa] from the end points of the continuum on a scale from 1 to 7 where1 = "very similar" and 7 = "very different." The test stimuli were presented in three different contexts: (1) same context with both sounds presented after the same vowel, (2) natural contexts with retroflex [ʂa] presented after [u] and alveolopalatal [ɕa] after [i], and (3) unnatural contexts with retroflex [ʂa] presented after [i] and alveolopalatal [ɕa] after [u].

**Results:** Responses were analyzed using mixed effects ordinal logistic regression models with subject as a random effect. An analysis with session (session 1,session 2) and context (same, natural, and unnatural) as fixed effects revealed a significant effect of session (LR.stat=40.56, df=1, $p<0.001$, the odds of rating the test stimuli as more dissimilar was 2.16 times higher in session 2 than in session 1), and interaction between session and context (LR.stat=16.46, df=2, $p<0.001$). To understand the nature of the interaction, we did separate analyses for each session. The session 1 analysis revealed no significant effect of context, suggesting that the perception of the novel sounds was not significantly dependent on information from the contexts before exposure. The session 2 analysis revealed a significant effect of context (LR.stats=20.64, df=2, $p<0.001$, the odds of rating the test stimuli as more dissimilar was 2.21 times higher in same contexts than in natural and unnatural contexts, and 1.59 times higher in unnatural contexts than in natural contexts). Participants' perception of the novel sounds became significantly dependent on information from the contexts after exposure. Of particular interest is that participants perceived the test stimuli as being more dissimilar from each other in the unnatural contexts than the natural contexts. This suggests that participants likely learned to do contextual cue integration after exposure in our previous study.

References:

[1]Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. Perception & Psychophysics, 28(5).

[2]Yu, A. C. (2011). On measuring phonetic precursor robustness: a response to Moreton. Phonology, 28(03).

[3]Anonymous (2014) Learning phonetic categories with phonotactics:the influence of predictability and phonetic naturalness. Poster presented at The 40th Annual Boston University Conference On Language Development.

[4]Maye, J. C. (2000). Learning speech sound categories from statistical information. Ph.D. dissertation, University of Arizona.