Within-category cue sensitivity in native language perception and its relation to non-native phonological contrast learning

Jieun Lee[1] & Hanyong Park[2]

[1]*University of Kansas (USA),* [2]*University of Wisconsin-Milwaukee (USA)*
jieunlee@ku.edu, park27@uwm.edu

Previous research has challenged assumption of categorical perception and provided evidence of gradient encoding of speech categories [1]. Previous studies with alternative measures of speech perception, such as 4IAX and the Visual Analogue Scaling tasks, have consistently shown listeners' ability to discriminate speech stimuli within the same category and a wide range of individual differences in their sensitivity to subtle within-category acoustic differences [2, 3, 4]. The current study attempted to quantify individual variability in within-category acoustic cue sensitivity in the perception of the native language (L1) category and its relation to second language (L2) phonological contrast learning (Experiment 1). We also investigated the effectiveness of Cue-Attention Switching Training (CAST). CAST is designed to downweight and upweight the irrelevant and the relevant acoustic cues of a target L2 phonological contrast, respectively, in native category perception. We examined whether such training would decrease the possible L2 learning gap due to the listeners' different sensitivities to the acoustic cues (Experiment 2).

Experiment 1: 24 Korean adult learners of English participated in learning three English vowel contrasts /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/ -/u/, which are distinguished primarily by spectral properties and secondarily by duration. The AXB oddity task was adopted to quantify participants' sensitivity to within-category differences induced by spectral and duration cue changes. A set of stimuli was constructed from the Korean /i/ vowel, with five different steps of spectral and duration properties. A and B stimuli in AXB pairs had either a one-step spectral or a one-step duration difference from the X stimulus, while the other step remained the same as the X stimulus. Participants were asked to judge whether A or B was more distinct from X and pick the "most odd one" out. We hypothesized that participants who dominantly selected stimuli with different spectral steps (i.e., High Sensitive (HS) group) would master the distinctions between English vowels of target contrasts better than those who did not show such high sensitivity to spectral steps (i.e., Low Sensitive (LS) group). In a pretest-posttest design, both groups received a five-day computer-based auditory L2 training with a set of stimuli comprising minimal pairs from the endpoints of a five-step spectral and duration continuum of either /i/-/ɪ/, /ɛ/-/æ/, or /ʊ/ -/u/: hid/heed, head/had, and hood/who'd. Overall, the HS group demonstrated an initial advantage in L2 contrast learning. This group showed higher scores in tests taken after each training day and more nativelike and increased use of spectral cues in the post-test compared to the LS group (Figs. 1 & 2). The individual-level analysis showed that despite individual variability in developmental trajectories in spectral and duration cue weightings over time, most HS group participants presented a desired trajectory pattern of enhancing the reliance on spectral dimension after the training. There were also positive correlations between participants' AXB oddity task results and their spectral cue weights for the target /i/-/ɪ/ and /ʊ/-/u/ contrasts, indicating that the higher selections of spectrally different stimuli are associated with the higher use of spectral dimension after the training (Fig. 3).

Experiment 2: A new group of 25 participants (LS2 group) took part in the same procedure as in Experiment 1 with short, simple two-alternative forced choice CAST before each training session. CAST stimuli were a subset of the AXB oddity task stimuli. We included stimuli varying along all five duration steps but only at the two most extreme spectral steps (steps 1 & 5). Participants identified each stimulus as either "Type 1 Korean /i/" or "Type 2 Korean /i/" and received feedback. The spectral properties of stimuli always determined the correct answer: "Type 1 Korean /i/" for stimuli with spectral step one and "Type 2 Korean /i/" for stimuli with spectral step five. The primary purpose of CAST was to downweight duration (i.e., make it an uninformative dimension for participants' decisions) and reallocate participants' attention to spectral cues. We observed some benefits of CAST, especially for the /i/-/ɪ/ contrast (Fig. 4). The

LS2 group identified the target English vowel contrasts better than the LS group in Experiment 1. Notably, the LS2 group exhibited more native-like use of the spectral cue than the LS group.

Our study demonstrated that L2 learners' individual differences in sensitivity to sub-phonemic acoustic details in L1 could explain variability in the learning of novel nonnative phonological contrasts. This suggests the transfer of L1 cue sensitivity to L2 cue utilization: how successfully L2 learners progress to become more nativelike listeners can be predicted in terms of their sensitivity to the L2 informative acoustic cue in L1 speech perception. Furthermore, we have shown that the addition of CAST can be helpful for L2 learners to overcome initial disadvantages due to such individual differences in L1 perception.

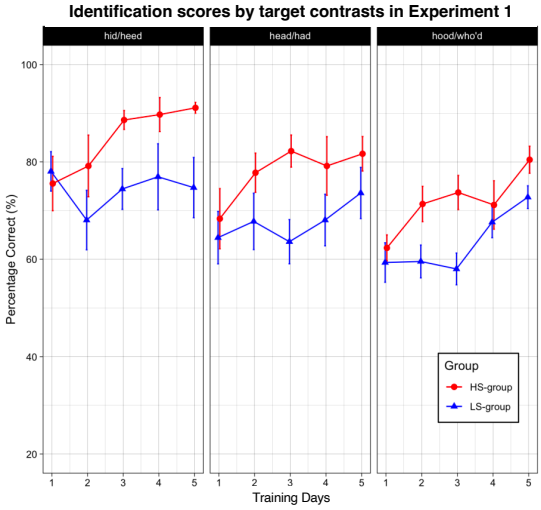

Fig.1 Identification test scores (% correct) after each training day by target contrasts for the HS-group (red) and the LS-group (blue) in Experiment 1.

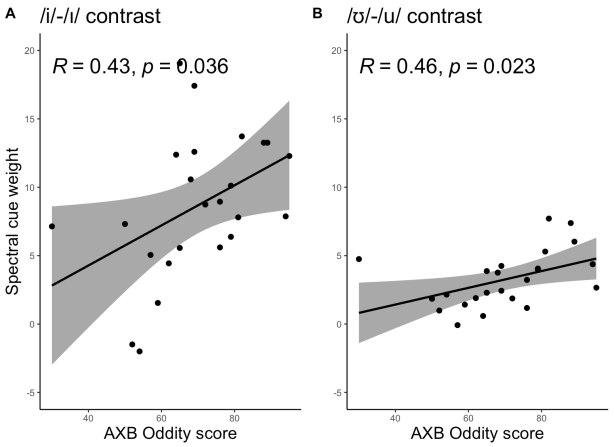

Fig.3 Correlations between individual participants' AXB oddity task scores and their beta-coefficients of spectral dimension for the /i/-/ɪ/ (left) and /ʊ/-/u/ (right) contrasts from the logistic regression analysis fitted to each participants' response data of the post-test.

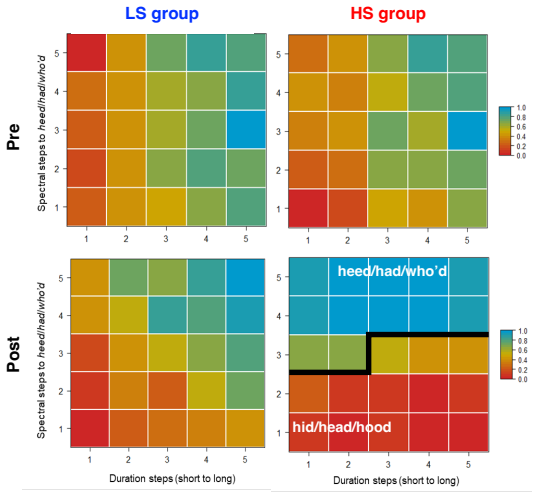

Fig. 2 Heat plots of overall responses of the pretest and the post-test in Experiment 1 to each combination of spectral and duration dimensions averaged across three vowel contrasts. The darkest red cells elicited 100% '*hid/head/hood'* responses, while the darkest blue cells elicited 100% '*heed/had/who'd*.'

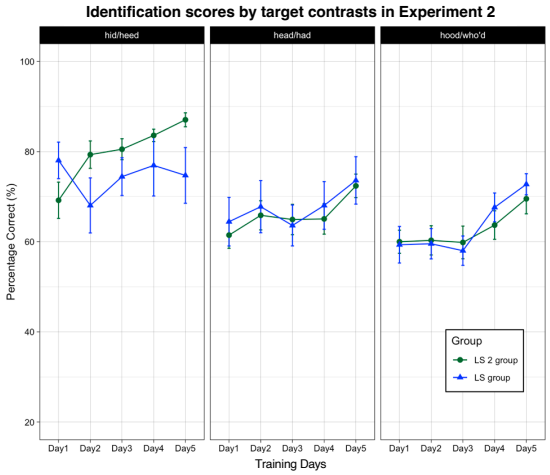

Fig. 4 Identification test scores (% correct) after each training day by target contrasts for the LS 2 group (green) and the LS-group (blue).

References

[1] McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819–3842.

[2] Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728–3745.

[3] Kong, E. J. & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

[4] Pisoni, D. B. & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328–333.