LEARNING TO DISTINGUISH THE THREE-WAY LARYNGEAL CONTRAST OF KOREAN PLOSIVES BY NATIVE ENGLISH SPEAKERS

Tyler K. Perrachione[1,2], John D.E. Gabrieli[2], & Amy S. Finn[2,3]

[1]Department of Speech, Language, and Hearing Sciences, Boston University, USA
[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA
[3]Department of Psychology, University of Toronto, Canada
tkp@bu.edu

ABSTRACT

Second-language acquisition often entails learning to use unfamiliar speech sounds to distinguish words, which may require learning to perceive an unfamiliar phonetic feature or accommodate a new category boundary along a known phonetic continuum. Here, native English speakers (*N*=37) learned an 18-word vocabulary comprised of six minimal triplets based on the Korean three-way plosive contrast (fortis, lenis, aspirated), defined by a trading relation between two phonetic features (onset pitch and voice onset time) that conflicts with how these cues structure listeners' native categories. Variation in these features across talkers and place of articulation further obfuscates this contrast. Learning outcomes were highly variable: Mixture model analysis suggested two groups of learners who differed primarily in whether they learned to distinguish lenis from aspirated stops. Both groups learned these contrasts best for bilabials and least accurately for alveolar stops. These results underscore the challenge of overcoming native category structure in second-language learning.

Keywords: second language acquisition, voice onset time

1. INTRODUCTION

A key challenge in acquiring a second language in adulthood is learning to perceive and produce novel phonological categories. Listeners begin to commit their perceptual and neural pathways to processing the sound structure of their native language during the first years of life [14], and thereafter it becomes increasingly challenging to learn to perceive or produce contrasts from other languages [2].

Two prominent factors constrain speech-sound learning in adulthood: *(i)* The extent to which a novel contrast's category boundary conflicts with a native boundary on the same phonetic continuum (e.g., stop categories with different voice onset time (VOT) boundaries, or vowels with different distributions in the F1×F2 space) [4], and (*ii*) the extent to which a novel contrast requires attending to a phonetic feature that is not contrastive in one's native language (e.g., F3 in the case of Japanese learners of English /r/-/l/ [13], or pitch contour direction in the case of English learners of Mandarin lexical tones [5]).

An anecdotally challenging contrast for native English speakers is the Korean three-way laryngeal contrast for onset plosives. This contrast distinguishes words like /ppul/ 뿔 "horn"; /bul/ 불 "fire"; and /pul/ 풀 "grass." In word onset, the Korean three-way stop consonant contrast can be characterized via a trading relation between three levels of positive VOT (short for fortis vs. intermediate for lenis vs. long for aspirated sounds) and onset f_0 (low for lenis vs. higher for fortis and aspirated sounds) (Fig. 1).

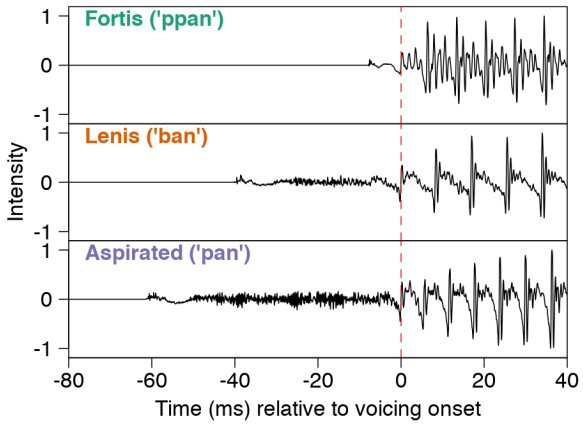

Figure 1: Acoustics of Korean plosive onset contrasts, aligned to voicing onset (dashed line). Note differences in both VOT and onset glottal period (f_0) across categories.

While naturally occurring minimal triplets are rare in Korean, there are multitudinous instances of both two-way minimal contrasts. Phonetic descriptions of these contrasts have defied simple characterization as a function of only voice onset time (VOT) or aspiration [1], and debate about the articulatory distinctions among the three contrasts [7] has led to inconsistencies in their transcription via the IPA. A straightforward articulatory or acoustic classification is further complicated by allophonic variation in the realization of these stops (particularly the lenis contrast) as a function of syllable position [16] and by variation in their realization across dialects [15]. To avoid mechanistic claims about articulation implicit in IPA transcriptions, in this report we use the standard Revised

Romanization transliterations: *fortis* /pp/, /tt/, /kk/; *lenis* /b/, /d/, /g/; and *aspirated* /p/, /t/, /k/.

Cross-linguistic work has historically focused on Korean learners of English [20], but the rapid global ascent of Korean pop culture has catalysed new interest in learning this language [24]. However, little empirical work has examined the interesting challenges that the fortis-lenis-aspirated contrast poses for English learners of Korean: First, its division of the VOT continuum into two aspirated categories conflicts with English listeners' expectation that aspiration reliably signals a voiceless (vs. voiced) word-initial stop. Second, onset stops in English (like many languages) covary in VOT and onset f_0, such that onset f_0 is low for voiced sounds like /b/ and high for voiceless sounds like /p/ [10,12]. This VOT-onset f_0 pattern is violated by Korean lenis stops, where a voiceless, slightly aspirated stop is realized with low onset f_0.

It is unclear how English learners will cope with the VOT-onset f_0 trading relation in Korean stops. Some have speculated that the fortis-lenis contrast will be hardest for English speakers, since the intermediate VOT and low onset f_0 of lenis stops may be confusable for English voiced stops [18]. Others have suggested that the lenis-aspirated contrast will be the most challenging [19], because aspiration unambiguously signals voiceless stops in English. Furthermore, despite the relationship between VOT and onset f_0 in English stops, English listeners are notoriously poor at learning to use pitch in phonological contrasts [5,22,23]. However, the atypically low onset f_0 of the lenis stop provides a correlated cue that may instead help listeners overcome their native VOT boundary.

2. METHODS

In this study we trained native-English speakers with no prior Korean experience to use the three-way Korean plosive onset contrast to identify a vocabulary of pseudoword triplets based on those contrasts.

2.1. Participants

Native English-speaking monolingual adults ($N = 37$, 25 female, 12 male, age 18-33, mean = 23.1 years) completed this study. Participants had minimal foreign language experience: none had more than two years of study, no earlier than in college. No participant had any prior experience with Korean. This study was approved by the IRB (COUHES) at the Massachusetts Institute of Technology.

2.2. Stimuli

Training stimuli consisted of 18 monosyllabic Korean pseudowords (Table 1). These words were organized into six minimal triplets (e.g., 'ppan', 'ban', 'pan'). There were two triplets per place of articulation (bilabial, alveolar, and velar). Familiar vowel and coda consonants were chosen to reduce learning demands.

빤 ppan (cow)	똑 ttok (bell)	깻 kkaet (car)
반 ban (seashell)	독 dok (sock)	갯 gaet (parrot)
판 pan (hammer)	톡 tok (grapes)	캣 kaet (camera)
삠 ppim (bus)	뗍 tteop (brush)	꿍 kkung (chair)
빔 bim (lamp)	덥 deop (box)	궁 gung (hat)
핌 pim (desk)	텁 teop (fish)	쿵 kung (fork)

Table 1: Vocabulary of minimal triplet Korean pseudowords (and target images).

High quality sound recordings were obtained from four adult native Korean speakers (two male, two female) of the Seoul dialect. Stimuli were RMS amplitude normalized to 65 dBA. As a control, two additional native Korean speakers confirmed the naturalness and distinctiveness of these stimuli and achieved perfect accuracy on the vocabulary training.

2.3. Acoustic measurements

We measured the key phonetic features of the three-way contrast (VOT and onset f_0) for all tokens in the vocabulary to validate the contrastiveness of our talkers' phonetic categories. VOT was measured as the latency between the release burst and the onset of phonation (the point on the waveform with the first high-amplitude, periodic deflection from zero). Onset f_0 was measured over the first five glottal cycles.

Each talker produced clear distinctions among the contrasts (Fig. 1, left). However, overlap between categories was also evident due to variation in mean f_0 and idiosyncratic VOT across talkers. In addition to being a critical feature of this contrast, VOT also varies for stops as a function of place of articulation, with longer VOTs for more posterior constrictions [1,18].

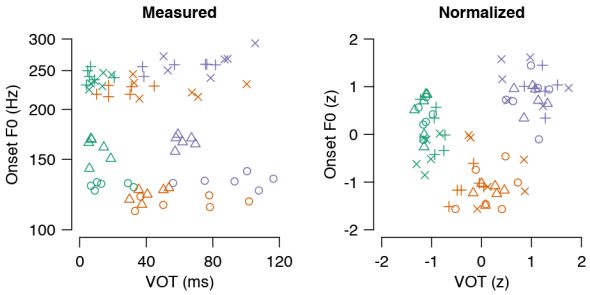

Figure 2: Phonetics of the vocabulary. Acoustics (left) show the challenge of overlapping category boundaries across talker and place of articulation (normalized at right). Symbols denote each talker, colours denote each contrast (green: fortis, orange: lenis, purple: aspirated).

A linear mixed-effects model of laryngeal contrast and place of articulation effects on VOT (by-talker random intercepts) revealed significant effects of laryngeal contrast ($F_{2,60} = 132.7$, $p < 0.0001$; aspirated > lenis > fortis) and place of articulation ($F_{2,60} = 20.8$, $p < 0.0001$; velar > labial > alveolar), but their interaction was marginal ($F_{4,60} = 2.20$, $p = 0.08$).

The corresponding model for onset f_0 showed a significant effect of laryngeal contrast ($F_{2,60} = 56.7$, $p < 0.0001$; aspirated > fortis > lenis) but not place of articulation ($F_{2,60} = 1.51$, $p = 0.23$) nor their interaction ($F_{4,60} = 0.45$, $p = 0.77$).

Normalizing VOT and onset f_0 within talker and place of articulation clearly reveals the three-way category structure in our vocabulary (Fig. 2, right).

2.4. Training procedure

Participants learned to associate each of the 18 Korean pseudowords with the photograph of a unique, familiar, natural object (Table 1). Images were shown in isolation on a white background. Participants completed four days of training, each consisting of a training and a test phase (Fig. 3). During the training phase, participants passively listened to the word-object pairings and actively practiced matching spoken words to the corresponding objects while receiving corrective feedback. During the test phase, participants actively matched spoken words to the corresponding objects without feedback [23].

Stimulus delivery was controlled using PsychoPy [21]. Participants completed each day's training and testing by themselves in a quiet room. Audio stimuli were delivered via Sennheiser HD-202 headphones at a comfortable volume selected by the participant.

During the training phase, participants completed interleaved blocks of passive familiarization and active practice. Words were trained in minimal triplets to emphasize the target contrasts. During passive familiarization, participants heard each word while the corresponding object was shown for 2000 ms (Fig. 3, left). Participants were instructed to attend to the stimuli and remember each word-object pairing. Each familiarization block contained 24 trials (3 triplet words × 4 talkers × 2 repetitions). Recordings from all four talkers were used during training, as high-variability training contributes to category learning [11], but stimuli were blocked by talker to reduce processing costs incurred by talker changes [9,22].

Participants next actively practiced matching the words in the triplet to the corresponding picture (Fig. 3, middle). While all three objects were displayed on the screen, participants heard each of the 12 recordings (3 words × 4 talkers) in a random order. Their task was to click on the corresponding picture using a mouse. Feedback was given immediately indicating whether they had chosen correctly or what the correct answer should have been. This self-paced procedure was repeated for a total of 24 active practice trials.

The interleaved familiarization-practice cycle of the training phase was repeated six times, once for each of the minimal triplets, in a random order. Thus, participants underwent 144 passive familiarization trials and 144 active practice with feedback each day.

After completing the training phase, participants' vocabulary learning was assessed in the Test Phase (Fig. 3, right). In this phase, an array of all 18 objects was shown. Participants heard all 72 recordings (18 words × 4 talkers) in a random order, selecting the corresponding object by clicking on it with a mouse. The test phase was self-paced and no feedback was given. Performance on the daily test phase was used as the dependent variable operationalizing learning.

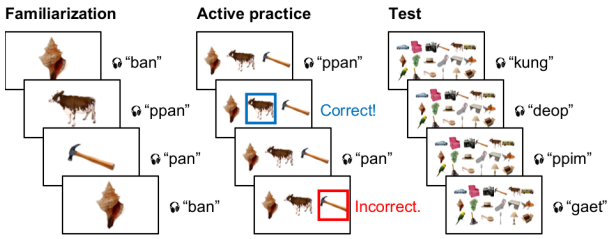

Figure 3: Vocabulary training procedure. Participants completed interleaved familiarization and active practice blocks for each triplet, followed by a test of all 18 items.

3. RESULTS

All participants improved with training, but learning outcomes on Day 4 were highly variable, with word-identification accuracy ranging from 22% to 69%. Mixture-model analysis of Day 4 accuracy indicated learning outcomes were best described by two subpopulations (Fig. 4): low-performance learners (LPL) comprised the bottom two-thirds of participants with average word learning of 37% ± 7%, while high-performance learners (HPL) comprised the top third of learners with average word learning of 63% ± 5%.

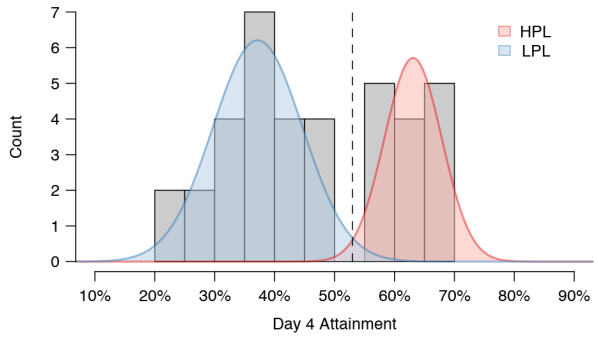

Figure 4: Mixture-model analysis revealed that learning outcomes were best captured by two distributions, one for high- and one for low-performing learners (HPL, LPL).

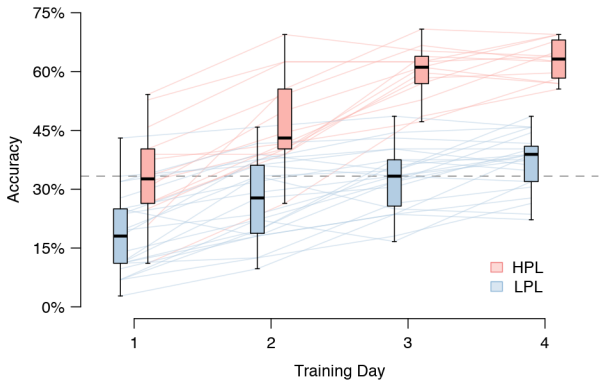

Figure 5: Learning progress by group. Lines show individual learning trajectories. Chance performance is shown for distinguishing between minimal triplets (1/3).

The HPL group learned faster than the LPL group even on Day 1, when they had already mastered the basic vocabulary triplets and were constrained by within-triplet phonetic contrast (chance level = 1/3), whereas the LPL group did not master between-triplet word-object associations until Day 4 (Fig. 5).

Patterns of word identification confusion can reveal the development of learners' novel phonological category structures. Plotting the patterns of confusion on Day 4 (Fig. 6) revealed that the HPL group had largely mastered the fortis contrast (83% ± 12% correct identification of fortis contrast words) and was making progress towards being able to use the lenis (55% ± 12%) and aspirated (51% ± 11%) contrasts to distinguish words. The LPL group, however, had not yet mastered using the fortis contrast to recognize words (50% ± 17% word identification) and was essentially at chance for identifying words distinguished by the lenis (32% ± 11%) and aspirated (30% ± 15%) contrasts. In general, both groups found the fortis stop most distinctive and had the most difficulty distinguishing lenis from aspirated onsets.

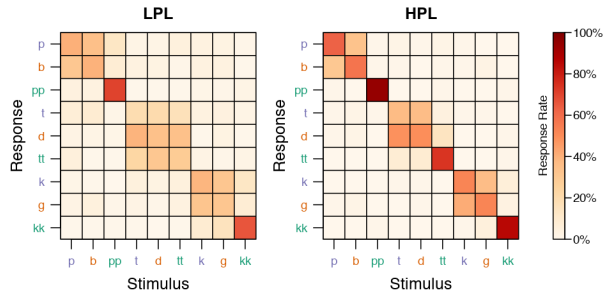

Figure 6: Onset consonant confusion matrices by group. More accurate learning of phonetic categories is captured by greater response clustering along the diagonal.

Both groups were most accurate at identifying words based on the three-way contrast involving bilabial stops (HPL: 70%, LPL: 44%), then velar stops (HPL: 63%, LPL: 44%), and were least accurate for alveolar stops (HPL: 57%, LPL: 22%).

4. DISCUSSION

Learning the Korean three-way laryngeal contrast for onset plosives is particularly challenging for native English speakers because (*i*) it divides a familiar phonetic continuum (VOT) in a way that conflicts with native voicing contrasts [1,4,18] and (*ii*) it violates the expected relationship between VOT and onset f_0 [10] in a way that requires listeners to pay attention to a perceptually challenging phonetic feature [6,17,23].

Participants varied widely in mastering the vocabulary involving these contrasts after four days of training. About one in three learners mastered word identification based on the fortis contrast (which is most similar to English /b/, /d/, and /g/) but still made mistakes identifying words based on the lenis vs. aspirated contrasts (which are both perceptually similar to English /p/, /t/, and /k/ in word onset). This pattern of learning is consistent with the prominence of aspiration in English voiceless stops [1,8], which appears to dominate perception even when onset f_0 provides a potentially disambiguating cue. Despite the VOT-onset f_0 relationship in English [10], it is not clear that listeners actually use this cue to perceive word-initial stops, and they may need to be explicitly instructed to attend to pitch in order to learn to use this feature [6].

Interestingly, the LPL group did not master even the fortis stop, despite its similarity to English /b/. It may be that partial overlap in the VOT distributions of the fortis and lenis stops across talkers led listeners to discount the contribution of this otherwise reliable cue, yielding a haphazard pattern of responses.

Also unexpected is learners' difficulty with the alveolar stops. On Day 4, even the HPL group was confusing fortis /tt/ for aspirated /t/ nearly as often as lenis /d/. This may be due to the unfamiliar pattern of VOT-by-place of articulation in this vocabulary: Whereas English onset stops tend to have increasing VOT with more posterior place of articulation, here alveolar stops' VOT was shorter than bilabial stops (lenis: /d/ 32 ms vs. /b/ 41 ms; aspirated: /t/ 59 ms vs. /p/ 74 ms). Combined with an unfamiliar amount of within-talker variation in VOT (resulting from lenis vs. aspirated contrast), these unexpected patterns of VOT may have upended English-speaking listeners' ability to generalize systematicity in VOT across talkers [8], further undermining their learning.

Overall, these results provide new empirical data to reaffirm that a core tenet of second-language speech-sound learning holds true even when a secondary cue is available to reinforce learning: The more that a novel phonological contrast conflicts with native categories, the harder it will be to learn [4].

5. REFERENCES

[1] Abramson, A.S. & Whalen, D.H. 2017. Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *J. Phonetics* 63, 75-86.

[2] Baese-Berk, M., Chandrasekaran, B., & Roark, C.L. 2022. The nature of non-native speech sound representations. *J. Acoust. Soc. Am.* 152, 3025-3034.

[3] Barcroft, J. & Sommers, M.S. 2005. Effects of acoustic variability on second language vocabulary learning. *Stud. Second Lang. Acq.* 27, 387-414.

[4] Best, C.T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words* 167, 233-277.

[5] Chandrasekaran, B., Sampath, P.D., & Wong, P.C.M. 2010. Individual variability in cue-weighting and lexical tone learning. *J. Acoust. Soc. Am.* 128, 456-465.

[6] Chandrasekaran, B., Yi, H.-G., Smayada, K.E., & Maddox, W.T. 2016. Effect of explicit dimensional instruction on speech category learning. *Attention, Perception, & Psychophysics* 78, 566-582.

[7] Cho, T., Jun, S.-A., & Ladefoged, P. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *J. Phonetics* 30, 193-228.

[8] Chodroff, E., Godfrey, J., Khudanpur, S., & Wilson, C. 2015. Structured variability in acoustic realization: A corpus study of voice onset time in American English stops. *18th International Congress of Phonetic Sciences* (Glasgow, August 2015).

[9] Choi, J.Y., Hu, E.R., & Perrachione, T.K. 2018. Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attention, Perception, & Psychophysics* 80, 784-797.

[10] Dmitrieva, O., Llanos, F., Shultz, A.A., & Francis, A.L. 2015. Phonological status, not voice onset-time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English. *J. Phonetics* 49, 77-95.

[11] Flege, J.E. 1995. Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguist.* 16, 425-442.

[12] Haggard, M., Ambler, S., & Callow, M. 1970. Pitch as a voicing cue. *J. Acoust. Soc. Am.* 47, 613-619.

[13] Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Deisch, E., Tohkura, Y., Ketterman, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.

[14] Kuhl, P.K. 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831-843.

[15] Lee, H., Politzer-Ahles, S., & Jongman, A. 2013. Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *J. Phonetics* 41, 117-132.

[16] Lee, H.B. 1993. Korean. *J. Int. Phon. Assoc.* 23, 28-31.

[17] Lim, S.-J., & Holt, L.H. 2011. Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science* 35, 1390-1405.

[18] Lisker, L. & Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.

[19] Oh, E. 2011. Effects of speaker gender on voice onset time in Korean stops. *J. Phonetics* 39, 59-67.

[20] Park, H. & de Jong, K.J. 2008. Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *J. Phonetics* 36, 704-723.

[21] Peirce, J.W. 2007. PsychoPy - Psychophysics software in Python. *J. Neurosci. Meth.* 162, 8-13.

[22] Perrachione, T.K., Lee, J., Ha, L.Y.Y., & Wong, P.C.M. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *J. Acoust. Soc. Am.* 130, 461-472.

[23] Wong, P.C.M. & Perrachione, T.K. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguist.* 28, 565-585.

[24] Zabel, S. 2021. "Squid Game" could inspire a new wave of Korean language learners. Available online: https://blog.duolingo.com/squid-game-could-inspire-a-new-wave-of-korean-language-learners/