THE PERCEPTION OF KOREAN STOPS BY NATIVE SPEAKERS OF SPANISH

María Teresa Martínez García[1] & Jeffrey J. Holliday[2]

[1]Hankuk University of Foreign Studies, [2]Korea University
[1]mtmg87@gmail.com, [2]holliday@korea.ac.kr

ABSTRACT

Whereas Spanish word-initial stops can be pre-voiced or voiceless, Korean word-initial stops are always voiceless, its three-way contrast being distinguished partly by short- vs. long-lag VOT. In intervocalic position, Spanish voiced stops become spirantized and Korean lenis stops become voiced, resulting in complicated phonological mappings among the stops in the two languages. In this exploratory study we tested the perceptual assimilation and discrimination of Korean stops by native Spanish (1) naïve listeners and (2) L2 learners of Korean.

It was found that both groups of listeners predominantly (roughly 70-90%) assimilated all Korean stops to a Spanish voiceless stop category, with the only major exception being intervocalic lenis stops. Differences between the listener groups were minor in the perception of word-initial stops, but much clearer in the case of intervocalic stops, which we attribute to effects of L2 experience. Discrimination accuracy was largely predicted by perceptual assimilation patterns.

Keywords: non-native speech perception; Spanish stops; Korean stops; perceptual assimilation

1. INTRODUCTION

The literature on the second language (L2) acquisition of non-native stop contrasts suggests that many errors in production are due to differences in the use of voice onset time (VOT) in the realization of the native (L1) and L2 stop contrasts [8, 9, 20, 22, 27]. It has separately been claimed that the mispronunciation of certain non-native sounds by L2 learners is due to L1 and L2 phonological categories existing in a shared phonological space, with certain L2 perceptual targets failing to be sufficiently differentiated from the targets of similar L1 sounds [7]. It follows that a central question in the acquisition of L2 stop systems is how the members of the L2 stop contrast are perceived, both by naïve listeners and L2 learners.

The Korean stop system contrasts tense (fortis) /p*, t*, k*/, lax (lenis) /p, t, k/, and aspirated /p^h, t^h, k^h/ stops. In word-initial position, VOT and fundamental frequency (f0) are the primary acoustic cues to the contrast: tense stops are cued by short-lag VOT and high f0, lax stops by long-lag VOT and low f0, and aspirated stops by long-lag VOT and high f0 [14, 15]. Previous studies have investigated the L2 acquisition of the Korean stop contrast by L1 speakers of English [19] and Mandarin [11], both languages with a short-lag vs. long-lag VOT stop contrast, and also Japanese [13, 26], a language with a voiced vs. voiceless stop contrast that has been claimed to be undergoing a sound change in which historically voiced stops are becoming short-lag voiceless stops [23, 24]. These studies found that Korean tense stops were assimilated by naïve English and Mandarin listeners to their L1 short-lag stop category, but by naïve Japanese listeners as to their L1 voiced stop category. It was speculated that the sound change in progress might have led Korean tense stops to be perceived as post-sound change Japanese voiced stops.

How would Korean stops be perceived by speakers of a language whose voiced stops are unequivocally pre-voiced? Spanish contrasts voiced (/b, d, g/) and short-lag voiceless (/p, t, k/) stops in word-initial position. These voiced stops are fully pre-voiced, with the onset of laryngeal vibration preceding the closure release by 40 ms or more [5, 10, 16]. If VOT is the primary cue by which L1 Spanish listeners perceptually assimilate non-native stops, it is predicted that naïve Spanish listeners would assimilate *all* word-initial Korean stops – tense, lax, and aspirated – to Spanish voiceless stop categories. In intervocalic position, Korean lax stops do become fully voiced [b, d, g], but Spanish voiced stops are realized as homorganic approximants [β̞, ð̞, ɣ̞] [10]. It is thus unclear whether any Korean stop in any position would ever be assimilated to a Spanish voiced stop category.

The current study aimed to explore the perceptual mappings of non-native Korean stops onto native Spanish categories. To generate hypotheses about L2 acquisition, we tested the perceptual assimilation and discrimination of Korean stops by both naïve Spanish listeners and native Spanish L2 learners of Korean. Comparing the perception of these two populations can shed light on how knowledge of the phonological system of Korean can affect acoustic perception (cf. [12, 21].

2. METHOD

A pre-registered study plan and R code are available at https://osf.io/afdj7/.

2.1. Participants

15 naïve Spanish listeners (tested in Alicante, Spain) and 15 L1 Spanish L2 learners of Korean (tested in Seoul, South Korea) participated in this study. The L1 Spanish L2 learners of Korean were all undergraduate students or professors at a university in Seoul. None of them started learning Korean before the age of 16, and they varied in length of residence in Korea (M = 26.9 months, SD = 20.6) and Korean instruction (M = 12.0 months, SD = 10.8).

2.2. Materials

The stimulus set consisted of Korean isolated CVs and VCVs produced by female native speakers of Seoul Korean. There were 216 CVs (3 phonation types × 3 places of articulation × 3 vowel contexts /a, i, u/ × 4 talkers × 2 tokens), and 72 VCVs (3 phonation types × 3 places of articulation x 1 vowel context /a/ × 4 talkers × 2 tokens).

2.3. Procedure

The experiment was presented on a notebook computer running OpenSesame ver. 3.2.5 [17]. Listeners completed five tasks in the following order: (1) CV perceptual assimilation, (2) VCV perceptual assimilation, (3) relative voicing perception, (4) CV discrimination, and (5) VCV discrimination. All written and verbal instructions were given in Spanish. Task order was not counterbalanced as it was not expected to have a significant impact on the results.

Perceptual assimilation: Participants heard one stimulus at a time and clicked on a button on the screen to indicate which Spanish stop the C sounded most similar to, given six choices presented orthographically: <b>, <p>, <d>, <t>, <g>, <k>. The CV block was presented before the VCV block, but stimuli were randomized within each block. All 216 CV and 72 VCV Korean stimuli were played once.

Relative voicing perception: Two CV stimuli produced by a single talker at the same place of articulation but of differing phonation types were played consecutively with an ISI of 750 ms. Participants then chose which of the two stimuli was more similar to the homorganic Spanish voiced stop. For example, if a trial consisted of Korean /pa/-/p^ha/, the participant would choose whether the first or second stimulus sounded more similar to Spanish /b/. A total of 108 trials were presented: 3 places × 3 vowel contexts × 3 contrasts (tense-lax, tense-aspirated, lax-aspirated) × 2 talkers × 2 orders.

Discrimination: Participants were presented with AXB trials of Korean stops, with an ISI of 750 ms. Each of the three stimuli in each trial was produced by a different talker, and the order of talkers was the same for every trial. There were 108 CV trials (3 places × 3 vowel contexts × 3 contrasts × 4 orders: AAB, ABB, BBA, BAA), and 36 VCV trials (same as above but with only one vowel context, /a/).

2.4. Data analysis

As this study is exploratory in nature, for the perceptual assimilation task we report simply the proportion of voiced and voiceless responses for each stimulus category and listener group. For the relative voicing perception task, the proportions of each Korean stop category that were perceived as more similar to a Spanish voiced stop category will be reported. Lastly, for the discrimination task we ran a repeated measures ANOVA to determine whether each listener group's accuracy varied by phonation type contrast, followed by planned comparisons among the AXB stimuli orders (cf. [6, 18, 25]).

3. RESULTS

3.1. Perceptual assimilation

In the CV trials, listeners from both groups assimilated all three Korean stop categories to a Spanish voiceless category roughly 70 to 90% of the time, as shown in the left side of Table 1. At least two general observations can be made from these results.

First, given the voiceless nature of word-initial Korean stops, it was expected that close to 100% of the stimuli would assimilate to Spanish voiceless categories, especially in the case of long-lag VOT lax and aspirated stops. An investigation of items and listeners showed that in both listener groups 5 of the 15 listeners accounted for over 70% of lax and aspirated stop trials assimilated to a voiced stop category, suggesting large individual differences in the perceptual assimilation of Korean stops. A similar level of inter-listener variation was observed in the responses to tense stops, with individual rates of assimilating tense stops to a voiced category ranging from 0 to 80% A thorough exposition of these individual differences cannot be presented here due to space limitations, but consequences of this finding will be discussed below.

Second, the L2 listeners' assimilation patterns were slightly more skewed toward Spanish voiced categories than the naïve listeners' were. Substantial inter-listener variation was observed here as well.

Table 1: Percentage (%) of each stimulus category assimilated to Spanish stop categories.

	CV			VCV		
Naïve	Ten	Lax	Asp	Ten	Lax	Asp
/b, d, g/	31.9	13.4	11.2	16.9	**55.6**	3.6
/p, t, k/	**68.1**	**86.6**	**88.8**	**83.1**	44.4	**96.4**
L2	Ten	Lax	Asp	Ten	Lax	Asp
/b, d, g/	31.6	17.6	10.2	9.2	**93.1**	3.1
/p, t, k/	**68.4**	**82.4**	**89.8**	**90.8**	6.9	**96.9**

In the VCV trials, a much larger difference between the listener groups was observed, as shown in the right side of Table 1. While Korean lax stops were perceived as both voiced and voiceless in roughly equal proportions by the naïve listeners, the L2 listeners were far more likely to perceive them as voiced (93.1%). Here as well, however, a high level of inter-listener variation was observed in the perception of lax stops by naïve listeners: although the mean assimilation rate to Spanish voiced stop categories was 55.6%, the individual rates ranged from 0 to 100%. These high levels of individual variation were not expected.

3.2. Relative voicing perception

The question addressed in this task was which Korean stop categories, when directly compared to each other, sound more like Spanish voiced stops. The results from both listener groups suggest that Korean tense stops are perceived as the most similar to Spanish voiced stops, but when lax stops are directly compared to aspirated stops, it is lax stops that sound moderately more similar to Spanish voiced stops, as shown in Table 2.

Table 2: Percentage (%) of trials in which each Korean stop category was chosen as more similar to a Spanish voiced stop category.

Naïve	Ten	Lax	Asp
Ten-Lax	**74.4**	25.6	
Ten-Asp	**73.7**		26.3
Lax-Asp		**55.2**	44.8
L2	Ten	Lax	Asp
Ten-Lax	**64.1**	35.9	
Ten-Asp	**78.1**		21.9
Lax-Asp		**62.8**	37.2

The differences between the naïve and L2 listeners were minor, but together suggest that L2 listeners perceive lax stops as slightly more similar to Spanish voiced stops than naïve listeners do. The tense-over-lax rate decreased from 74.4% to 64.1%, and the lax-over-aspirated rate increased from 55.2% to 62.8%.

3.3. Discrimination

The results of the CV discrimination trials are shown in Table 3. The *accuracy* column is the overall accuracy for the contrast in each row. The *A→B* and *B→A* columns break down the overall accuracy into trials that that were AXB and BXA, respectively, (e.g. in the first row, which is tense-lax, A→B includes /p*a/-/p*a/-/pa/ and /p*a/-/pa/-/pa/, whereas B→A includes /pa/-/p*a/-/p*a/ and /pa/-/pa/-/p*a/).

Table 3: Discrimination accuracy (%) and d′ on CV trials. A→B includes AAB and ABB trials, whereas B→A includes BAA and BBA trials.

Naïve	accuracy (d′)	A→B	B→A
Ten-Lax	67.0 (0.88)	61.9 (0.60)	72.2 (1.19)
Ten-Asp	67.2 (0.90)	61.1 (0.56)	73.3 (1.26)
Lax-Asp	53.3 (0.17)	53.0 (0.16)	53.7 (0.18)
L2	accuracy (d′)	A→B	B→A
Ten-Lax	63.1 (0.67)	60.7 (0.54)	65.6 (0.79)
Ten-Asp	63.0 (0.66)	62.2 (0.62)	63.7 (0.70)
Lax-Asp	52.0 (0.10)	53.3 (0.18)	50.7 (0.04)

Overall, the CV discrimination results reveal differences in accuracy among the contrasts in both naïve [$F(2,28) = 14.4$, $p < .001$] and L2 listeners [$F(2,28) = 9.5$, $p < .001$], with planned comparisons showing that listeners are more accurate at discriminating tense stops from lax and aspirated stops than they are at discriminating lax and aspirated stops from each other (all comparisons $p < .01$). With respect to stimulus order, planned paired t-tests revealed a significant effect of order for naïve listeners in tense-lax [$t(14) = -3.3$, $p = .005$] and tense-aspirated [$t(14) = -3.2$, $p = .007$] trials, such that listeners were more accurate when first stimulus was lax or aspirated. No effect was found for naïve listeners on lax-aspirated trials, or for L2 listeners at all.

Table 4: Discrimination accuracy (%) and d′ on VCV trials. A→B and B→A are used as in Table 3.

Naïve	accuracy (d′)	A→B	B→A
Ten-Lax	79.4 (1.61)	85.6 (2.02)	73.3 (1.20)
Ten-Asp	58.3 (0.42)	50.0 (0.00)	66.7 (0.83)
Lax-Asp	82.8 (1.91)	81.1 (1.71)	84.4 (2.09)
L2	accuracy (d′)	A→B	B→A
Ten-Lax	81.7 (1.78)	80.0 (1.62)	83.3 (1.90)
Ten-Asp	56.1 (0.30)	55.6 (0.27)	56.7 (0.33)
Lax-Asp	83.9 (1.94)	82.2 (1.78)	85.6 (2.02)

The results of the VCV discrimination trials are shown in Table 4. Both naïve [$F(2,28) = 26.1$, $p < .001$] and L2 [$F(2,28) = 20.1$, $p < .001$] listeners exhibited differences in accuracy among the three contrasts, with planned paired t-tests revealing significant differences between the tense-aspirated

trials and the two other contrasts for both listener groups (all comparisons $p < .001$). With respect to stimulus order, planned paired t-tests revealed a significant effect for naïve listeners in tense-aspirated [$t(14) = -2.2$, $p = .046$] trials, such that listeners were more accurate when first stimulus was aspirated. No other order effects were found.

Taken together, the results from the discrimination tasks are largely predictable from the perceptual assimilation patterns reported in Section 3.1.

4. DISCUSSION

Given that word-initial Korean stops are always voiceless, it was initially predicted that naïve Spanish listeners would assimilate all word-initial Korean stops to a Spanish voiceless stop category. It was found that they often did, but not always, and our exploration of why revealed a high degree of inter-listener variability. This result was unexpected, as most other studies of the perceptual assimilation of Korean stops reported a high degree of consistency across naïve listeners [11, 13, 19, 26], presumably due to perceptual assimilation being influenced by the acoustic properties of the stimuli alone. It is less surprising that inter-listener variability would be found among L2 listeners, as their perceptual assimilation would be affected not only by the acoustic properties of the stimuli, but also the specific properties of the phonological representations of Korean stop categories they have developed so far, which could easily differ across individuals.

Because our original study plan did not include detailed comparisons between individual listeners, we instead lay out here our plans for ongoing analyses. First, it is possible that listeners might be modifying their response strategy over the course of the experiment: responses to early trials might favour a voiceless response, but as no truly voiced stimuli are presented listeners might start to shift their responses to adapt to the VOT scale presented in Korean stops. In an experimental manipulation described in the pre-registration but for which data collection is still ongoing, we have mixed in native Spanish stop productions among the Korean stop stimuli in the perceptual assimilation task to see whether the presence of truly voiced stimuli would bias listeners to assimilate the Korean stimuli to Spanish voiceless categories.

Second, we plan to investigate whether different listeners are weighting acoustic cues differently in the Korean stop stimuli. In the most extreme example of inter-listener variability, two naïve listeners assimilated intervocalic lax stops to Spanish voiced categories 0% of the time, while one listener did so 100% of the time. Given that Spanish intervocalic voiced stops are spirantized, it seems that some listeners weigh the lack of spirantization in Korean more than the presence of stop voicing, resulting in such stimuli being assimilated to a Spanish voiceless category. Although we cannot explicitly test cue weighting in the current dataset, we believe the current results can be used to generate hypotheses for a follow-up study. The issue of individual differences in stop perception by L1 Spanish listeners has been discussed in the literature [1, 2], but not examined systematically.

Third, the range of responses among the L2 listeners warrants a closer look at their individual linguistic backgrounds. Because some of the L2 learners had much more experience with Korean than others, it is possible that some of the inter-listener variation, among the L2 listeners at least, could be partly explained by their level of L2 experience.

Lastly, our interpretation of the discrimination results was based on group-level patterns, and it was indeed shown that in both the CV and VCV trials, both groups of listeners were in aggregate more accurate at discriminating between Korean stops that were assimilated to different categories, in line with predictions made in the literature [3, 4]. But because the aggregate group results were actually reflections of wide inter-listener variability, it is clear that our evaluation of the discrimination results should be done at the level of individual listeners, including our analysis of stimulus order effects.

In [25], discrimination accuracy in an AX task was more difficult when the stimuli proceeded from a well-assimilated stimulus to a poorly assimilated stimulus. In the current study, at the group level, word-initial Korean aspirated stops assimilated better to voiceless stops than tense stops did, suggesting that discrimination accuracy should be poorer when an aspirated stop is followed by a tense stop. The opposite trend was found here in both the tense-aspirated and tense-lax contrasts (see Table 3). But this result was found only in the naïve listeners, and it does not consider whether the listeners who exhibited the largest asymmetry in Table 3 are the same listeners who assimilated both aspirated and tense stops to the same Spanish category. Thus, there are numerous individual listener comparisons that should be done in order to form more specific testable hypotheses for subsequent experiments.

5. ACKNOWLEDGEMENTS

We thank professor Tae-Yeoub Jang for his assistance with recording the stimuli, and Valentín Martínez García and María Teresa García Seller for their help with data collection in Spain.

6. REFERENCES

[1] Abramson, A. S., Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

[2] Alves, U. K., & Luchini, P. L. (2016). Perception of the distinction between voiceless and voiced initial stops in English by Argentine learners: Identification and discrimination data. *Lingüística, 32*(1), 25-39.

[3] Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

[4] Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

[5] Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*(2), 207-229.

[6] Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 270-282.

[7] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

[8] Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*(3), 185–202.

[9] Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech, 24*(2), 125–146.

[10] Harris, J. (1967). *Spanish phonology* (Unpublished dissertation). Cambridge, Mass.: MIT Press.

[11] Holliday, J. J. (2014). The perceptual assimilation of Korean obstruents by native Mandarin listeners. *Journal of the Acoustical Society of America, 135*(3), 1585–1595.

[12] Holliday, J. J. (2016). Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica, 73*(1), 33–51.

[13] Holliday, J. J. (2018). The perception and production of word-initial Korean stops by native speakers of Japanese. *Language and Speech*.

[14] Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics, 39*(2), 196–211.

[15] Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association, 42*(2), 145–169.

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

[17] Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

[18] Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 221-231.

[19] Schmidt, A. M. (2007). Cross-language consonant identification: English and Korean. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 185–200). Amasterdam, The Netherlands: John Benjamins.

[20] Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2015). Effects of age and speaking rate on voice onset time. *Studies in Second Language Acquisition, 37*(1), 71–100.

[21] Sturman, H. W., Baker-Smemoe, W., Carreño, S., & Miller, B. B. (2016). Learning the Marshallese phonological system: The role of cross-language similarity on the perception and production of secondary articulations. *Language and Speech, 59*(4), 462–487.

[22] Swanson, K. A. B. (2006). *Acquisition versus suppression of phonological processes in the second language acquisition of French and English.* Ph.D. Dissertation. Bloomington: Indiana University.

[23] Takada, M. (2008). VOT variation in Japanese word-initial stops. In B. Heselwood & C. Upton (Eds.), *Proceedings of methods XIII: papers from the Thirteenth International Conference on Methods in Dialectology* (pp. 372–382). Frankfurt am Main, Germany: Peter Lang.

[24] Takada, M., Kong, E. J., Yoneyama, K., & Beckman, M. E. (2014). Age- and gender-related variation in voiced stop prenasalization in Japanese. *Journal of the Acoustical Society of America, 136*(4), 2174–2174.

[25] Tsushima, T., Shiraki, S., Yoshida, K., & Sasaki, M. (2003). On stimulus order effects in discrimination of nonnative consonant contrasts. *Acoustical Science and Technology, 24*(6), 410-412.

[26] Yasuta, T. (2004). *Stop perception in second language phonology: Perception of English and Korean stops by Japanese speakers.* Ph.D. Thesis. University of Hawai`i.

[27] Zampini, M. L. (1998). The relationship between the production and perception of L2 Spanish stops. Texas *Papers in Foreign Language Education, 3*(3), 85–100.