PERCEPTUAL ASSIMILATION OF SPANISH, THAI, AND KOREAN STOP CONSONANTS BY NATIVE ENGLISH SPEAKERS

Charlie Nagle[1], Shelby Bruun[1], Melissa Baese-Berk[2]

The University of Texas at Austin[1], University of Oregon[2]
cnagle@austin.utexas.edu, sbruun@utexas.edu, mbaesebe@uoregon.edu

ABSTRACT

Models of second language (L2) sound learning argue that learners' initial perceptual assimilation patterns are important because they shape perceptual learning pathways and outcomes. To contribute to this line of literature, we examined native English listeners' perceptual assimilation of Spanish, Thai, and Korean stop consonants, asking participants to map stops from all three languages onto English stop categories and rate their goodness-of-fit using a 5-point scale.

Results showed that several categories were unequivocally assimilated to a single English category (e.g., Korean Aspirated and Lenis stops to English voiceless stops). In contrast, the Spanish Voiceless category, Thai Plain category, and Korean Fortis category were mapped onto both English voiced and voiceless stops. Yet, this group trend masks substantial individual variation in the assimilation of these ambiguous categories. Some participants showed split categorization, aligning with the group, whereas others showed a strong preference for one of the English categories.

Keywords: second language, speech perception, perceptual assimilation, stop consonants, individual differences

1. INTRODUCTION

During the first year of life, infants can easily perceive differences between sounds, including sounds that do not occur in their native language (L1) [1]. However, as L1 categories coalesce, the ability to distinguish between non-native (L2) sound contrasts rapidly diminishes. L1 speech learning "warps" perceptual space, such that the L1 acts as a perceptual filter for the L2 [2]. The individual no longer perceives acoustic reality but rather their perception is pushed and pulled by L1 categories.

Models of L2 speech sound learning such as the Speech Learning Model [3, 4] and the Second Language Perceptual Assimilation Model [5] formalize how the L1 and L2 should interact during learning and perception. For example, the Speech Learning Model argues that the timing of L2 learning relative to L1 learning is critical. For individuals who learn the L2 later in life, by the time L2 learning begins the L1 system is already firmly in place, which means that it will exert a strong influence on the L2. Likewise, according to the L2 Perceptual Assimilation Model [5], initial perceptual assimilation patterns determine the ease or difficulty of the learning task. For instance, if both members of an L2 contrast are assimilated to a single L1 category and judged to be relatively equally good exemplars of that category, then learning to perceive the L2 contrast could be very challenging. Despite their differences, the Speech Learning Model and the L2 Perceptual Assimilation Model share the assumption that crosslinguistic assimilation patterns matter, insofar as they may dictate both the starting point and the rate of L2 perceptual learning.

Additionally, there is likely to be substantial individual variation in these patterns, even among listeners who come from the same L1 background. The L1 sets the boundaries on what is possible in terms of potential L2 assimilation patterns, but within those general, language-specific guardrails, individual listeners appear to exhibit a range of meaningful subpatterns [6]. Such findings mirror the extensive individual variability documented in other areas of perceptual performance and learning, including individual differences in phonetic cue weights [7, 8].

Stop consonants are one of the most studied L2 segments. Surprisingly, however, there have not been many perceptual assimilation studies targeting L2 stops. This may be because stop consonant systems have a smaller set of categories than vowel systems, where multicategory mappings (i.e., many-to-one or many-to-many) are common (e.g., [6]). In other words, most languages have two to three stop consonant categories at a given place of articulation, which places an inherent limit on the types of crosslinguistic patterns that may emerge. Yet, even if patterns are relatively fixed for L2 stops, it is important to understand how strong those patterns are, that is, how well the L2 sound fits into the L1 category, which could have an impact on learning new L2 stop categories.

The few studies that have been carried out on the perceptual assimilation of word-initial stops show that listeners tend to respond based on the phonetic cues present in their L1. For instance, Mandarin [9], Japanese [10], and Thai listeners [11] assimilate

Korean stops to the L1 category that is the best match in terms of aspiration, or the amount of voice onset time (VOT) present in the stop, presumably because VOT is the primary cue to stop consonant contrasts in each language. Thus, these listeners tend to show a single category assimilation for Korean Lenis and Aspirated stops, which are similar in terms of VOT but differ with respect to fundamental frequency (F0), while making a distinction between those categories and Korean fortis stops, which are produced with shorter VOT. Oliveira and Rato [12] reported similar findings for Cantonese listeners' perception of European Portuguese stops. It bears mentioning that these trends may mask substantial variability in individual response patterns [10].

Further, current research has focused on the perceptual assimilation of stops from one additional language by listeners who are unfamiliar with that language. Investigating how individuals assimilate sounds from several languages would shed light on the relative difficulty listeners might experience when learning those languages. It would also be interesting to examine individual patterns of variation to document the range of assimilation patterns and the strength of those patterns within languages and within listeners.

In this study we therefore examined English listeners' perceptual assimilation of word-initial stops in Spanish, Thai, and Korean. Briefly stated, English and Spanish have two phonological stop consonant categories, whereas Thai and Korean have three. In all languages, VOT is a relevant phonetic cue to stop consonant identity, and it serves as the primary cue in English, Spanish, and Thai [13, 14]. In Korean, however, both VOT and F0 are essential cues to the three-way contrast, which cannot be fully differentiated by either cue independent of the other [8, 15]. Importantly, although English speakers are sensitive to F0 differences, they primarily rely on VOT, which is a more robust cue to stop consonant voicing in English [16].

2. METHOD

We created the experiment using Gorilla Experiment Builder (www.gorilla.sc) and integrated it with Prolific (www.prolific.co) for participant recruitment. Access was restricted to desktop devices located in the United States.

2.1. Participants

Twenty-eight monolingual speakers of English (13 females, 15 males) completed the study. Mean participant age was 35.39 *(SD* = 11.82). 89% (*n* = 25) of participants identified as White, 4% (*n* = 1) as

Black, and 7% (*n* = 2) as Mixed or Other. All participants were born and raised in the United States with one participant reporting a period of residency in the United Kingdom. All reported normal hearing.

2.2. Tasks

Participants completed a background questionnaire targeting known languages and language use. None reported proficiency in the target languages.

Participants then completed a forced-choice perceptual assimilation task including 320 items, described below. The instructions made it clear that all stimuli would begin with either a "p" sound or a "b" sound. Participants were asked to indicate the sound they heard by selecting one of two choices: "Sounds like English 'p'" or "Sounds like English 'b'". Goodness ratings were elicited following each trial on a 5-point scale where only the endpoints were labelled (1 = "bad" and 5 = "good"). The task was completed in three blocks (2 blocks of 107 trials, 1 block of 106 trials) to allow for two two-minute rest periods between blocks. Tokens from each language were intermixed across the task and presentation order was randomized across participants.

2.2.1 Perceptual Assimilation Stimuli

We used a subset of stimuli from a larger stimulus set containing Thai, Korean, and Spanish words and nonsense words. All nonsense words obeyed the phonotactic constraints of their respective language. To validate the stimuli, we asked native speakers of each language to evaluate how good of an example each stimulus was of the intended word on a 4-point scale (1 = "poor", 4 = "excellent"). All stimuli used in this experiment were rated 3 or higher.

In this study, there were eight stop consonant categories: two for Spanish and three each for Thai and Korean. Within each language, the stimuli were produced by four talkers, two males and two females, and the initial stops were combined with five following vowels for two repetitions (4 talkers × 5 vowels × 2 repetitions = 40 stimuli per category). Thus, there were 320 total stimuli.

2.3.1. Acoustic characteristics of stimuli

Stimuli were coded for VOT and F0 using Praat version 6.1.53 and values were extracted using a script. Positive VOT was measured from the release of the stop consonant burst to the onset of voicing of the following vowel. Negative VOT was measured from the onset of vocal fold vibration, evident as low-frequency periodicity in the waveform, to the burst of the stop consonant. Ten percent of the data (50/320

tokens) was cross-coded to ensure intercoder reliability. Following Schertz et al. [8], F0 was measured 5 ms after the onset of the following vowel. Table 1 summarizes the acoustic characteristics of the stops in each language.

	VOT	F0
	M (SD)	M (SD)
Spanish		
Voiced	−104.08 (33.31)	155.45 (16.99)
		139.05 (11.59)
Voiceless	22.68 (9.06)	182.35 (31.27)
		160.75 (28.78)
Thai		
Voiced	−77.08 (26.82)	244.25 (15.55)
		134.95 (5.52)
Plain	11.38 (4.29)	267.85 (14.54)
		155.90 (23.02)
Aspirated	91.26 (19.13)	271.50 (18.65)
		176.61 (30.02)
Korean		
Fortis	13.05 (5.23)	288.15 (31.78)
		150.50 (13.03)
Lenis	53.08 (15.67)	245.05 (14.00)
		124.95 (8.91)
Aspirated	70.72 (17.99)	326.10 (29.51)
		163.15 (11.22)

Table 1: Acoustic characteristics of the stimuli. F0 is shown separately for female (top) and male (bottom) talkers.

3. RESULTS

For each language, we examined the proportion of English voiced and voiceless stop responses per category and the corresponding category goodness rating. We computed proportions and mean goodness ratings for all participants to shed light on group trends. We also examined individual categorization patterns.

As shown in Figure 1, in Korean, the Aspirated and Lenis categories were mapped almost exclusively onto the English voiceless category and rated as good exemplars of that category. In contrast, the Fortis category was variably mapped onto both voiced and voiceless stops and received category goodness ratings that were highly variable. In Thai, the Aspirated category was mapped nearly exclusively onto English voiceless stops, whereas the Voiced category was mapped nearly exclusively onto English voiced stops, and both stops were deemed to be good exemplars of their English counterpart. Perception of the Plain category was more variable, though participants showed a slight preference for English voiced stops. Finally, for Spanish stops, the Voiced category was mapped onto English voiced stops, but

assimilation of the Voiceless category was variable. In this case, contrary to the pattern evident in Thai, Spanish Voiceless stops were generally mapped onto English voiceless stops.
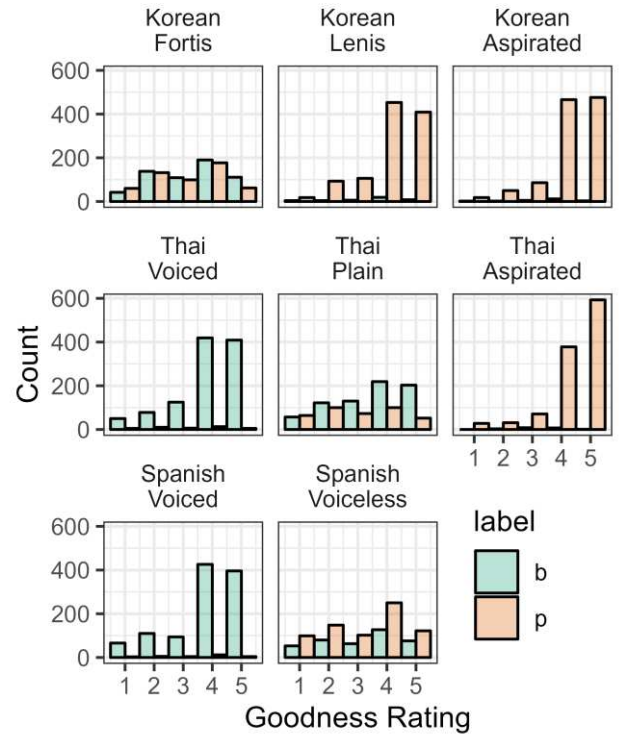

Figure 1: Assimilation patterns and goodness ratings for Korean, Thai, and Spanish stops.

We were particularly interested in examining assimilation patterns and goodness ratings for the categories that showed the most variability, namely Korean Fortis, Thai Plain, and Spanish Voiceless stops. To do so, we computed difference scores for each participant by subtracting the percentage of /b/ responses from the percentage of /p/ responses. Thus, in Figure 2 positive values index a preference for /p/, negative values a preference for /b/, and a value of zero no preference, or equal proportions of /b/ and /p/ responses. The plot reveals a wide range of patterns with respect to the direction of assimilation in each language. For instance, while nearly all individuals assimilated the Spanish Voiceless category to English /p/, for some participants this assimilation was relatively weak, near the zero mark, whereas for others it was quite strong at 75%. The opposite pattern is evident for Thai Plain stops, which most participants mapped onto English /b/ to varying degrees. This divergent pattern could be due to the slight difference in VOT across the languages, where Thai Plain stops were produced with shorter VOT than Spanish Voiceless stops (10 ms on average). Korean Fortis stops showed a relatively split pattern, with some participants mapping them onto English /b/ and others onto English /p/, and the difference scores

suggest a narrower range than what was observed for the other two languages. This narrower range seems to index greater ambiguity in terms of how the Korean Fortis category maps onto English stop categories, which could be due to a combination of VOT intermediate to the Spanish Voiceless and Thai Plain categories and higher F0, which would be associated with voicelessness in English.

Figure 2 also provides a visual summary of how distinctly each participant categorized the three ambiguous categories. Individuals on the extreme left and extreme right of the plot assimilated all three categories to the same English category (/b/ on the left and /p/ on the right), and for the rightmost participant on the plot, proportions were nearly identical across languages. Individuals shown in the middle portion of the plot tended to map Thai Plain stops onto /b/ and Spanish Voiceless stops onto /p/, showing the greatest uncertainty (i.e., values closest to zero) for Korean Fortis stops. Yet, even among these individuals the proportion of stops mapped onto one category varied considerably.
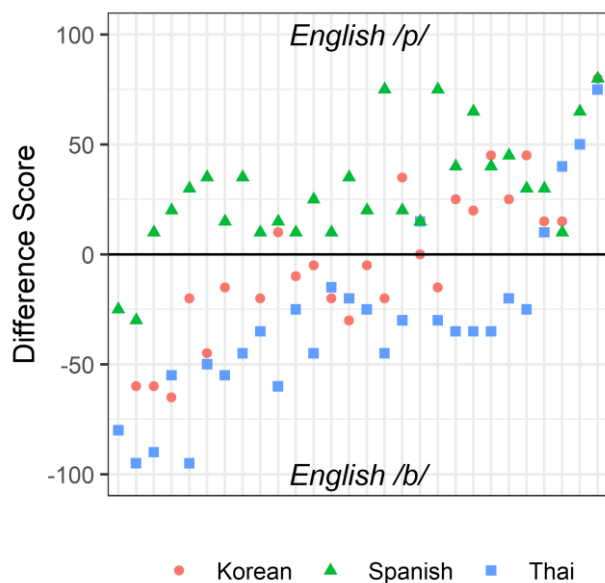

Figure 2: Individual assimilation patterns for Korean Fortis, Spanish Voiceless, and Thai Plain stops.

4. DISCUSSION

In this study, we examined perceptual assimilation of voicing contrasts from three languages by native English listeners. Each language showed slightly different assimilation patterns. Below, we review these assimilation patterns and the individual variability within each language and category type.

We found evidence for a single category assimilation for Korean Aspirated and Lenis stops, both of which were perceived as good exemplars of English voiceless stops. This finding mirrors the results of previous studies involving participants whose L1 predominantly marks stop consonant contrasts using VOT [9, 10, 11]. In other words, because those two Korean categories are both produced with a considerable amount of VOT, they are mapped onto the L1 category that is the closest phonetic match, which for English listeners is the voiceless category. In contrast, Korean Fortis stops were variably mapped onto English voiced and voiceless stops. The individual data revealed that most listeners mapped Korean Fortis stops onto English voiced stops, but there were several who mapped them onto voiceless stops. This means that for some listeners, all three Korean stops were mapped onto a single English category, albeit with varying degrees of fit, but the Lenis and Aspirated categories were always rated as better exemplars of English voicelessness than the Fortis category was.

For Thai, Voiced stops were mapped onto English voiced stops and Aspirated stops onto English voiceless stops, which fits with the distribution of VOT values in both languages. The Thai Plain category, however, was more ambiguous, insofar as it was mapped onto both English categories. Yet, most listeners perceived Thai Plain stops as a better exemplar of English voiced stops, at least in terms of response proportions. Thus, there seemed to be less ambiguity in the classification of Thai Plain stops than there was in the classification of Korean Fortis stops. In fact, some individuals classified Thai Plain stops as instances of English voiced stops nearly categorically. This also fits with the phonetic reality of English voiced stops, which can be realized either as fully (pre)voiced stops or as short-lag stops.

Finally, Spanish Voiced stops were mapped onto English voiced stops and rated as good exemplars of the English category, but the Spanish Voiceless category was more ambiguous. Most participants assimilated Spanish Voiceless stops to English voiceless stops, but the fit was variable.

Overall, these results provide baseline data for English listeners' perceptual assimilation of stops in three unfamiliar languages. They also highlight the importance of going beyond group trends to look at individual assimilation patterns [6]. The English listeners recruited for this study were mostly monolingual, reporting little to no L2 learning experience with the languages included in this study (or any other), and therefore represented a relatively linguistically homogenous participant sample. Yet, even these individuals showed varying perceptual assimilation patterns that could lead to different learning outcomes. Thus, based on these results, it is important to categorize individuals based on individual performance rather than group data.

6. REFERENCES

[1] Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9(2), F13–F21.

[2] Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B* 363, 979–1000.

[3] Flege, J. E. 1995. Second language speech learning: Theory, findings, problems. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross-Langauge Research*. York Press, 233–277.

[4] Flege, J. E., Bohn, O.-S. 2021. The revised Speech Learning Model. In: Wayland, R. (ed), *Second Langauge Speech Learning: Theoretical and Empirical Progress*. Cambridge University Press, 3–83.

[5] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, M. J., Bohn, O.-S. (eds), *Second Langauge Speech Learning: The Role of Language Experience in Speech Perception and Production*. John Benjamins, 13–24.

[6] Mayr, R., Escudero, P. 2010. Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition* 13(3), 279–297.

[7] Kim, D., Clayards, M., Goad, H. 2018. A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics* 67, 1–20.

[8] Schertz, J., Cho, T., Lotto, A., Warner, N. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52, 183–204.

[9] Holliday, J. J. 2014. The perceptual assimilation of Korean obstruents by native Mandarin Listeners. *The Journal of the Acoustical Society of America* 135, 1585.

[10] Holliday, J. J. 2019. The perception and production of word-initial Korean obstruents by native speakers of Japanese. *Language and Speech* 62(3), 494–508.

[11] Wayland, R. 2000. Perceptual assimilation and categorical discrimination of Korean stop consonants by native Thai speakers. *The Journal of the Acoustical Society of America* 108, 2652.

[12] Oliveira, D., Rato, A. 2018. Assimilação perceptiva das oclusivas orais do português europeu L2 por falantes nativos de cantonês. *Diacrítica* 32(2), 133–156.

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

[14] Raphael, L. 2021. Acoustic cues to the perception of segmental phonemes. In: Pardo, J. S., Nygaard, L. C., Remez, R. E., Pisoni, D. B. (eds), *The Handbook of Speech Perception*. Wiley Blackwell, 603–631.

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

[16] Shultz, A. A., Francis, A. L., Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132, EL95–101.