The influence of perceived L2 sound categories in on-line adaptation and implications for loanword phonology

Abstract
Some propose that loanword adaptation is at its core non-native perception of foreign input (Boersma and Hamann 2009; Peperkamp et al. 2008; Silverman 1992). It has also been noted, however, that cross-language correspondences in loanwords are far more consistent than expected based on on-line perception by naïve monolinguals. There is also evidence that cross-language perception itself differs depending on adapters’ experience with the source language (henceforth, L2) (Bundgaard-Nielsen et al. 2011; Kwon 2017; Nomura and Ishikawa 2016). These findings suggest that cross-language perception is mediated by adapters’ knowledge of L2 sound structure, rather than a simple function of native language (L1) perception applied to L2 acoustic signals. The current study presents a direct test of the influence of L1 vs. L2 perceptual strategies on cross-language speech perception through a series of phonetic categorization experiments in three language modes: L1, L2, and L2L1 (cross-language). Results point to a distinct influence of listener’s L2 knowledge on cross-language perception: L2L1 mapping was well explained by listeners’ L2 perceptual strategies, and for those listeners who showed different perceptual patterns for L1 and L2, cross-language perception more closely mirrored L2 than L1 perception. By demonstrating that perceived L2 phonological categories shape cross-language perception, the study suggests a way to reconcile the perceptual view of loanword adaptation with the phonological regularity of established loanwords.

Keywords: Loanwords; Cross-language perception; Korean; English

1 Introduction
1.1 Cross-language perception and loanwords
Studies on loanword phonology show that native phonological restrictions play a crucial role in reshaping foreign words in adaptation. Studies have also found, however, that there are many aspects of loanword adaptation that are not explained by the requirements of native phonology alone (see Kang (2011) for an overview). A considerable body of work has accumulated to show that such unexpected adaptation patterns are explained when the phonetic details of source language (henceforth, L2) and the native language (henceforth, L1) are taken into account. Specifically, it has been proposed that adapters select the L1 structure that is perceptually most similar to the L2 input as the optimal mapping (Boersma and Hamann 2009; Kang 2003; Kenstowicz 2007; Steriade 2001; Peperkamp et al. 2008; Silverman 1992; Yun 2016). For example, French coda nasal [n] is adapted with a paragogic vowel (and nasal gemination) in Japanese as in French Cannes [kan] > Japanese [kannu], even though *[kan] is a licit Japanese form. On the other hand, English coda nasal [n] is adapted without an extra vowel as in English pen > Japanese [pen], *[pennu]. The seemingly "unnecessary" adaptation in French loans is attributed to the phonetic realization of coda nasals in French: their long and intense vowel-like release is perceived by Japanese adapters as an extra vowel (Peperkamp et al. 2008).

Some analyses propose that adaptation happens during speech perception proper and that loanword adaptation is a direct function of native language perception applied to foreign input (Boersma and Hamann 2009; Peperkamp et al. 2008; Silverman 1992). While differing in details, these “adaptation-as-perception” models share an assumption that the input to adaptation (i.e., perception), is an unstructured phonetic signal devoid of any phonological information of the source language. This input is then filtered through the native perceptual decoder to arrive at the output (i.e., the phonological representation of the adapted loanword). In other words, loanword adaptation is equated with non-native perception by naïve listeners without any knowledge of source language sound structure, in the sense of PAM (Perceptual Assimilation Model) (Best 1995). This mapping is

* The authors would like to thank Dr. Sungwoo Han, Yunyan Luo, and Yuanyang Song for their assistance with data collection. The paper benefited from the comments from the associate editor, the reviewers, and the audience at the 2016 Annual Meeting on Phonology at USC and the Workshop on the Phonetics and Phonology in Loanword Adaptation held at Cologne in 2017. This research was supported by SSHRC Grant #435-2013-2092 to Yoonjung Kang.
schematically represented by the (blue) dotted arrow in Figure 1. Following Boersma and Hamann (2009), [] represents phonetic forms while // represents phonological forms in this figure.

Figure 1: Models of perceptual mapping in loanword adaptation; The blue dotted arrow represents cross-language mapping by a naïve listener without any knowledge of the source language (L2); The green dashed arrow represents a cross-language mapping via listeners’ perception/knowledge of source language phonological categories.

This adaptation-as-naïve-perception model of loanword adaptation is in contrast to the view that adapters are competent bilinguals with native-like knowledge of L2. Under this view, the phonological structure, not the phonetic forms, of the source language serves as input to the adaptation process (LaCharite & Paradis 2005). Yet others assume a more nuanced view of the role of source language phonology and recognize that the adapters’ L2 knowledge may not be native-like but nevertheless constitutes one of various factors that affect the outcome of adaptation (Kang 2008; Chang 2012; de Jong and Cho 2012; Ito 2014; Ito, et al. 2006; Kang 2010; Kenstowicz 2005; Kwon 2017; Smith 2006, 2009). Specifically, many propose that adaptation refers to subphonemic phonetic details of the source language, but that adapters’ source language knowledge also plays a role. Chang (2012) in particular proposes that the input to loanword adaptation is enriched to include adapters’ phonological knowledge of L2 while the adapters also attend to phonetic details of the L2. Smith (2009) proposes a model of loanword adaptation where the adapters’ “posited” source language phonological representation enters into a correspondence relationship with the output, and this posited phonological representation may be influenced not only by perceptual information, but also by explicit knowledge of the source language grammar and orthographical information.

Others propose a diachronic pathway for L2 knowledge to shape loanwords. Based on the observation that source-to-native correspondences in established loanwords are far more consistent than expected based on non-native perception by naïve monolinguals, Kang (2010) proposed that while perceptual similarity plays a primary role in shaping adaptation, the variation in loanwords is constrained by speakers’ knowledge of the phonological categories of the source language, which exerts a regularizing pressure. Kang (2010) envisions the adapter’s L2 knowledge as a grammatical constraint that overrides a perceptually faithful mapping, while the regularization itself can be seen as a diachronic process that operates on pre-existing loan variants. Similarly, de Jong and Cho (2012) note that while loanword mapping largely conforms to perceptual mapping, the two are not identical, and

---

1. See the figures in Silverman (1992, p. 293), Peperkamp, et al. (2008, p. 154), and Boersma and Hamann (2009) for the specific proposals from each model.

2. Note that Boersma and Haman (2009)’s use of brackets differ from the conventional usage of brackets in phonology; in addition to the phonetic form ([ ]) and the surface phonological forms (/ /), they assume an additional level of representation, i.e., an underlying lexical representation indicated by (| |). As our study is not concerned with phonological alternation where the surface phonological form and the underlying form may differ, we will not be concerned with the third level of representation.

3. The role of source language orthography on loan adaptation has been observed by many including Haugen (1950), Dohlus (2005), Vendelin and Peperkamp (2006), and Kang (2009, 2012).
mismatches between the two can be attributed to “historical lexicalization” and “explicit sociocultural standards” such as the standardization pressure by the National Institute of the Korean Language.

It has also been argued that the level of L2 knowledge at the community level affects the pattern of adaptation (Haugen 1950, Heffernan 2007). In his seminal work on linguistic borrowings, Haugen (1950) observes that sound adaptation evolves from “erratic substitution” to “systematic substitution,” in which “the same [native language] phoneme is consistently employed for new [input language] loans” as the community evolves from a “pre-bilingualism period” to an “adult bilingualism period”. Similarly, based on diachronic variation in Sino-Japanese loans, Heffernan (2007) observes that the higher the level of bilingualism, the more likely the adaptation will refer to phonological over phonetic representations of the input language.

While recognizing that L2 phonology can influence loanwords through more than one route, our study explores the influence of listener’s source language knowledge on cross-language perception proper. This study builds on the insight from recent findings that cross-language perception, while undeniably grounded in the acoustics of the input sounds, is modulated by listeners’ knowledge and experience with the input language. Specifically, Best and Tyler (2007) extended their model of nonnative perception (PAM) to second language (L2) perception (PAM-L2) and proposed that while naïve listeners perceive and process unfamiliar language sounds at the phonetic level, L2 listeners may process L2 sounds at the phonological as well as the phonetic level. Strange (2011)’s Automatic Selective Perception (ASP) model argues that while beginning learners of L2 employ a “phonetic mode of perception”, advanced L2 learners are predicted to develop “selective perception routines” that efficiently identify L2 phonological categories. Chang (2015) also proposes that “relatively experienced L2 users are expected to show L1-L2 mappings that follow phonemic similarity over acoustic and allophonic similarity because of a tendency for high-level information to override low-level information” (p. 204).

For example, using a typical paradigm to test cross-language perception, Bundgaard-Nielsen et al. (2011) asked L1 Japanese/L2 Australian English participants to listen to naturally produced Australian English (L2) vowels, then choose the Japanese (L1) vowel category that best matched the sound they had heard. Results showed that listeners with a larger L2 vocabulary selected fewer distinct L1 variants for a given L2 input vowel and showed more consistent cross-language categorization patterns than listeners with a smaller L2 vocabulary. They conjectured that L2 learners attune their perception of L2 sounds and establish L2 phonological categories through lexical acquisition, which in turn affects their cross-language mapping. Levy (2009) examined American English (AE) listeners’ mapping of French front rounded vowels, [y] and [œ], to English vowels and found that AE listeners with a high degree of L2 experience exhibited higher “internal consistency” — i.e., they were more likely to converge on the same AE vowel choice for a given French vowel — than listeners with less experience. In the study, the high experience group also showed less variation due to consonantal contexts (bilabial vs. alveolar), indicating that their mapping is more consistent vis-à-vis the L2 phonological category, abstracting away from context-dependent allophonic variation of perceptually similar L1 vowels.

The effect of L2 experience on cross-language perception has also been attested in illusory vowel perception (Dupoux et al. 1999). Nomura and Ishikawa (2016) examined Japanese listeners’ perception of an epenthetic vowel [u] in consonant clusters. In the experiment, the listeners heard English words containing a medial consonant cluster, illicit in Japanese, and pressed a button as soon as possible if they detected a target mora (for example, English word *homesick* [hɔmsɪk] had a target mora of á [ɔu]). The study found that intermediate-level English learners reported an illusory vowel less often than introductory-level learners, and based on this result, the authors proposed that higher-proficiency learners can switch their attention to L2 phonotactics more readily and inhibit L1-driven perception. Kwon (2017) examined Korean listeners’ perception of paragogic vowels following English word-final stops. Previous studies found that the presence of optional release in English word-final stops is equated with a vowel release by Korean adapters, because in Korean, final stops are obligatorily unreleased (Kang 2003). Kwon (2017)’s study found that when presented with nonce English words with a final stop, near-monolinguals (those who never lived in an English-speaking environment) and late bilinguals (those who learned English after the age of 10) chose a Korean output with a paragogic vowel more when the stimuli had an audible release burst than not, as expected if cross-language perception is a direct function of acoustic similarity. On the other hand, for early bilinguals, presence vs. absence of release had no effect on response
patterns, suggesting that in cross-language perception, higher-proficiency listeners abstract away from phonetic details (i.e., stop release) that are not relevant to phonological contrasts in the source language.\(^4\)

These studies suggest that cross-language perceptual assimilation can be mediated by listeners’ knowledge of source language phonological categories and that more proficient L2 listeners are better able to inhibit L1 influence, in turn relying more on their knowledge of the L2, when processing L2 sounds.\(^5\) As a result, adapters with some level of L2 knowledge may first map the L2 signal to L2 phonological structures through their L2 decoder (which is distinct from the native perceptual decoder), and these perceived L2 categories are then mapped to equivalent L1 categories. This L2-mediated route of perception is schematically represented by the green dashed line in Figure 1.

The studies discussed above have generated predictions about how L2 sounds may be mapped to L1 categories by examining how results of cross-language mapping tasks can be explained when taking into account knowledge of the distinct phonetic properties of L1 vs. L2 sounds. However, in these studies, the connection between cross-language mapping and perception of L1 and L2 is inferred rather than directly tested. The current study directly tests the assumption of congruence between L1 perception and cross-language perception (henceforth, L2L1 perception) that forms the basis of purely perceptual models of loanword adaptation (Silverman 1992, Boersma & Hamann 2011, and Peperkamp, et al. 2008). We further examine if and how perceived L2 categories mediate L2L1 perception, and whether this L2 influence varies depending on proficiency, using cross-language perception of English stops by Korean listeners as a test case.

1.2 Korean and English stops

English contrasts voiced and voiceless stops, and in word-initial position, aspiration (henceforth voice onset time, VOT) is the primary cue, while f0 (fundamental frequency, the acoustic correlate of pitch) and other secondary cues are also relevant. Korean contrasts three types of stops—aspirated [pʰ], lenis [p] and fortis [p’] —and word-initially they are distinguished primarily by VOT and f0 on the following vowel (see Cho et al. 2002 for a review). Aspirated stops are marked by a long VOT and a high f0, while lenis stops are marked by a mid-to-long VOT and a low f0. Fortis stops are marked by a short VOT and a mid-to-high f0.\(^6\) The distribution of Korean and English stops in the VOT-by-f0 acoustic space is illustrated by Figure 2. The English data are from Schertz (2014), and the Korean data are from participants in the current study.

---

\(^4\) It should also be noted that L2 knowledge does not necessarily lead to more faithful and consistent perception of L2 contrasts. For example, Holliday (2016) found that naive listeners were more accurate in Korean fricative discrimination (/s/ vs. /s’/) than novice learners of L2 Korean, suggesting that the explicit metalinguistic instruction and exposure to written language can override phonetically faithful perception.

\(^5\) Another possible explanation for why L2 experience might affect L2L1 mapping relies on the fact that L1 perception itself can change due to contact with the L2, and this in turn could affect loanword adaptation (Hamann and Li 2016). Under this view, loanword adaptation can still be conceptualized as L1-filtered perception of L2 input (as proposed in Boersma and Hamann 2009), but we still predict a congruence between L2 perception and L2L1 mapping. However, in this case, the effect of L2 experience on L2L1 mapping is indirect and occurs via a change in L1 perception proper.

\(^6\) Seoul Korean (along with many other dialects of Korean) is undergoing a sound change in which younger speakers produce a smaller VOT difference between lenis and aspirated stops than older listeners, instead using f0 as the primary cue for the contrast (Kang 2014; Silva 2006).
Previous studies show that English word-initial voiceless stops are mapped to Korean aspirated stops with high consistency and confidence by Korean listeners, while English voiced stops are variably mapped to Korean lenis or fortis stops (Park and de Jong 2008; Schmidt 1996). These results from cross-language perception experimental tasks are mirrored in established loanwords where English voiceless stops are consistently adapted as aspirate stops of Korean, while English voiced stops are variably adapted to lenis or fortis stops, as illustrated by examples in (1) (Kang 2008; Kim 2016; Oh 2009, 2017).

(1) English | Korean
--- | ---
pen | pʰen (English voiceless -> Korean aspirated)
belt | pɛl̩ɨ (English voiced -> Korean lenis)
bus | p̩as̩ɨ (English voiced -> Korean fortis)

The phonological correspondence between English voiceless and Korean aspirated stops is expected given the acoustic similarity between these two classes of sounds in Korean and English (phonetically voiceless, aspirated stops). At the same time, as shown in Figure 2, the categories are not acoustically identical: a naïve filtering of the English input through Korean categories in L2L1 mapping (the blue arrow in Figure 1) would predict English voiceless stops to be perceived sometimes as lenis and sometimes as aspirated Korean stops. On the other hand, if L2L1 mapping is mediated by L2 categories (the green arrow in Figure 1), then, assuming that English voiceless and Korean aspirated stops form phonological equivalence classes, we would predict a more “phonologically consistent” mapping of English voiceless stops to Korean aspirated stops. Therefore, this pair of sounds provides a good test case for the influence of L2 phonology in L2L1 mapping because the acoustic distribution of the two sounds is similar but not identical, generating distinct predictions for purely L1-based mapping vs. L2-mediated mapping. Furthermore, based on previous work on L2 perception of English by Korean listeners, we expect to find proficiency-conditioned variability in perception, with more proficient listeners showing more native-like perception of English (Kong & Yoon, 2013). This variability allows us to test our hypotheses about how L2 perceptual proficiency (which we will estimate based on proximity to native English perception) influences the extent of L2 phonological influence on L2L1 mapping. In other words, we expect that high-proficiency listeners will show a more consistent mapping of L2 to L1 phonological categories in cross-language perception.

In addition to having more native-like L2 perception, we hypothesize that more proficient L2 listeners will rely more heavily on their own L2 phonology in cross-language mapping. In order to test the relative influence of L1 and L2 on cross-language mapping, we conducted a series of perception experiments in which we charted Korean listeners’ perception of a parallel phonological contrast in their L1 (Korean) and their L2 (English), then compared the results to an L2L1 cross-language mapping task in which the same listeners heard L2 (English) sounds and chose the L1 (Korean) category that best matched what they heard. In order to test our proficiency-
based hypotheses, we quantified the Korean listeners’ perceptual proficiency by comparing their L2 results with a control group of native English listeners. In a departure from previous studies on cross-language perception, the use of controlled acoustic spaces for all of the tasks allows for direct comparison across the different modes of perception. We test two hypotheses which would follow from an L2-mediated model of L2L1 mapping, summarized in (2):

(2) Hypotheses and predictions following from an L2-mediated model of L2L1 mapping

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Expectations for the current work</th>
</tr>
</thead>
<tbody>
<tr>
<td>During cross-language perception, more proficient L2 listeners have a more consistent mapping between cross-language phonological correspondents than less proficient listeners.</td>
<td>More proficient listeners will show more consistent mapping of English voiceless stops to Korean aspirated stops in L2L1 perception, as compared with less proficient listeners.</td>
</tr>
<tr>
<td>More proficient listeners have a higher relative reliance on L2 (more reliance on L2 and/or more inhibition of L1) in cross-language perception than less proficient listeners.</td>
<td>More proficient listeners’ Korean aspirated stop responses in L2L1 mapping will more closely approximate their own L2 perception of English voiceless stops than their own L1 perception of Korean aspirated stops, as compared with less proficient listeners.</td>
</tr>
</tbody>
</table>

It is possible that cross-language mapping becomes more consistent for proficient listeners (hypothesis 1) without increased relative reliance on L2 over L1 (hypothesis 2). This would be consistent with a situation where more consistent cross-language mapping is achieved for proficient listeners due to the change in their L1 perception, which becomes more similar to L2 due to increased L2 influence (see footnote 5). If both hypotheses hold true, this provides support for direct mediation of listeners’ L2 in cross-language mapping. This L2-mediated model of L2L1 mapping can provide a solution to the paradox noted by Kang (2010) and de Jong & Cho (2012), namely, that loanword adaptation shows sensitivity to perceptual similarity between the source form and the native output, while at the same time, established loans are more consistent and systematic than naïve L1-based perceptual filtering would predict.

2 Methods

2.1 Experiments

Korean listeners completed three phoneme identification tasks summarized in Table 1. In Korean perception (L1), listeners heard Korean stimuli (sounds produced by a Korean speaker) and were asked to indicate which Korean sound they heard, while in English perception (L2), they heard English stimuli and responded with English categories. In cross-language perception (L2L1), a task intended to simulate perceptual adaptation, listeners heard English stimuli and responded with Korean categories. The instructions were in English for the L2 task and in Korean for the L1 and the L2L1 tasks. The stimuli for the two languages were manipulated to be identical in their range of VOT and f0 values, enabling us to directly compare the perceptual mapping across the three tasks in a controlled acoustic space. If cross-language perception is a function of L1 perception applied to foreign inputs, we predict that the L2L1 and L1 responses will be largely isomorphic. On the other hand, if cross-language mapping is guided by perceived L2 categories, L2L1 responses should mirror L2 responses. English listeners without any knowledge of Korean were also recruited to complete a task where they heard English stimuli and responded with English categories. This task is intended to establish the English perception norm, against which we evaluate the Korean participants’ L2 perceptual proficiency.

---

7 As an anonymous reviewer pointed out, L1 and L2L1 tasks shared the same response choices while L2 did not and in this sense the three tasks are not completely comparable. If anything, this should result in increased congruence between the L1 and L2L1 tasks; which would make it more difficult to find evidence for an L2-mediated model.
Table 1: Perception tasks by listener group

<table>
<thead>
<tr>
<th>Listener group</th>
<th>Task</th>
<th>Stimuli</th>
<th>Response choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>L1: Korean perception</td>
<td>Korean</td>
<td>Korean categories: (ㅍ/ㅃ/ㅉ)</td>
</tr>
<tr>
<td></td>
<td>L2: English perception</td>
<td>English</td>
<td>English categories: (ba/pa)</td>
</tr>
<tr>
<td></td>
<td>L2L1: Cross-language perception</td>
<td>English</td>
<td>Korean categories: (ㅍ/ㅃ/ㅉ)</td>
</tr>
<tr>
<td>English</td>
<td>English Control</td>
<td>English</td>
<td>English categories: (ba/pa)</td>
</tr>
</tbody>
</table>

2.2 Participants

88 Korean listeners (46 females and 42 males) born and raised in the Seoul metropolitan area completed the study and were paid for their time. 67 of the participants were recruited in the Seoul metropolitan area in the summer of 2015 and 21 in Toronto, Canada, in the spring of 2015. Four Seoul participants were excluded from analysis; one participant who chose a single response for all stimuli, two participants who showed random responses, and one whose L2 responses were the opposite of the expected direction (low VOT stimuli elicited voiceless responses).

All but two of the Seoul participants never resided in an English-speaking country for more than 6 months. Toronto participants varied in their length of residence in Canada: 11 of the participants were short-term visitors who recently arrived in Toronto to study English and had resided in Toronto for less than 6 months, while 10 had resided in Toronto between 7 and 13 years. A wider age range is represented by Seoul participants (year of birth: 1933-1996) than Toronto participants (year of birth: 1987-1995). The Toronto participants on average had higher L2 perceptual proficiency (as defined in section 3 below) than Seoul participants (Seoul: 65.9% vs. Toronto: 78.3%), but a wide range of proficiency is represented in both groups (Seoul: 36.3–93.7%, Toronto: 58.1–91.3%). For a subset of 31 participants recruited in Seoul, L2 production data, in the form of 32 short English sentences, was also collected. The recorded sentences were subsequently rated by four native speakers of English for accentedness (1: not accented ~ 5: heavily accented) and the speakers varied widely in average rating (range: 1.47–4.08; mean = 2.65).

We had no a priori predictions of differences between the Toronto and Seoul participants, apart from those differences that would follow directly from the differences in proficiency mentioned above. In initial exploratory data analysis, we confirmed that after controlling for L2 perception proficiency, Seoul and Toronto participants showed similar response patterns, and we therefore pool the data from the two groups in all of our analyses.

10 English listeners with no knowledge of Korean, recruited in Toronto in 2016, participated as a control group. Reflecting the demographics of Toronto, most English listeners were multilingual but they all identified English as their primary language. The languages the participants listed as their other language included Ukrainian, Cantonese, Mandarin, Bangla, Portuguese, and Arabic. Demographic information about the 84 Korean participants and 10 English control participants is provided in Table 2.

<table>
<thead>
<tr>
<th>Language</th>
<th>City</th>
<th>Decade of Birth</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>30s 40s 50s 60s 70s 80s 90s</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>Seoul</td>
<td>2M 2F 5M 5F 4M 7F 4M 9F 1M 1F 1M 2F 10M 7F</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Toronto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Toronto</td>
<td>1F 3M 6F</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2: Participant distribution by native language, recruitment location, decade of birth, and gender.

2.3 Stimuli

Parallel nonsense words in Korean [paɾu]/[pʰaɾu]/[p’aɾu] (바루/파루/빠루) and English [paɾu]/[baɾu], pronounced with initial stress, were recorded by one male native speaker of each language. Nonsense words with intervocalic rhotics were chosen to provide a stimulus-internal cue to the language; between vowels, the Korean liquid is realized as a rhotic tap while English has a rhotic alveolar approximant. Creation of stimuli was

8 Seoul data were collected as part of a larger study on sound change in Korean.
identical for Korean and English. We first created baseline tokens for manipulation by splicing aspiration—from the release burst of the stop to the onset of the periodic voicing of the following vowel—from an aspirated (Korean) or voiceless (English) token onto two vowels varying in voice quality. Voice quality, often quantified by H1-H2 (Cho et al., 2002), is an additional cue to the Korean stop contrast, and primarily distinguishes fortis from lenis/aspirated stops. We included baseline tokens with extreme values of H1-H2 (i.e., one fortis and one aspirated vowel for Korean) in order to control for this. Similarly, English stimuli were varied in the baseline vowel (i.e., one voiced and one voiceless vowel). However, unlike f0 and VOT, which could be controlled to be identical across the two language conditions, the base vowel difference across the two languages does not vary along the same acoustic dimensions or to the same degree. In order to keep the levels of the acoustic predictors identical across languages, we therefore did not include baseline vowel as a factor in our analyses. The different vowels serve to represent the range of natural variation present in each language.

These baseline tokens were equalized for duration and intensity in each language, then manipulated using the PSOLA algorithm in Praat (Boersma & Weenik 2016) to vary systematically in VOT (0-120ms, 8 steps) and f0 at vowel onset (83-120 Hz, 5 steps). In this way, we created a controlled “acoustic space” that was identical in the two languages, as shown in Figure 3. Each stimulus was presented twice, resulting in a total of 160 trials (2 baseline tokens * 8 VOT steps * 5 f0 steps * 2 repetitions) in each of the three tasks.

![Figure 3: Distribution of stimuli in the VOT-by-f0 acoustic space](image)

2.4 Procedure

Listeners heard the nonce-word stimuli and responded with the best-fit Korean (p/p’/pʰ) or English (p/b) category. For each task, instructions and response choices were presented in the “response” language (English for L2, Korean for L1 and L2L1), and listeners were specifically told that they would be hearing English (L2, L2L1) or Korean (L1) sounds. The L1 and L2 tasks were completed before the L2L1 task, and the order of L1 and L2 tasks was counterbalanced across participants. Participants heard the stimuli through headphones and responded with key presses on a MacBook Pro (Toronto) or touch screen taps on a Microsoft Surface Pro tablet (Seoul).

3 Results

Prior to analysis, Korean responses were converted to a binary choice of “aspirated vs. non-aspirated.” Fortis and lenis responses were collapsed to non-aspirated in Korean as they are both exponents of English voiced stops in loanword adaptation, and the use of a binary response variable allowed for a direct comparison with the L2 English “voiceless vs. voiced” responses. First, we examine the overall patterns of responses. Figure 4 shows the average rate of aspirated/voiceless responses (ASP RATE) aggregated over all Korean listeners for each of the three tasks—L1, L2L1, and L2. The figure also provides an average rate of voiceless responses by English control listeners for comparison (the rightmost panel). Darker shading indicates more aspirated or voiceless responses. As expected from the production patterns shown in Figure 2, there is a striking difference in distributions between Korean (L1) and English (control) responses; while in English the voiceless vs. voiced contrast is distinguished almost exclusively by VOT, the Korean aspirated vs. non-aspirated contrast is distinguished by a combination of VOT and f0. Figure 5 provides another view of the same observation by
plotting the average aspirated/voiceless responses by VOT conditions separately for the lowest (83Hz) and the highest (120Hz) f0 conditions. Korean listeners are sensitive to VOT—the lines have a positive slope—and also to f0—the solid and the dashed lines are well separated. On the other hand, English listeners rely mostly on VOT, while the value of f0 makes little difference. Another observation to draw from these figures is that the Korean listeners’ responses are similar across all three tasks. Since Korean L1 perception is markedly different from native English perception, we can therefore conclude that Koreans’ L2 and L2L1 perception are both strongly influenced by their L1 perception.

![Figure 4: Average aspirated/voiceless response rates for each task by Korean listeners and English control listeners](image)

![Figure 5: Average aspirated/voiceless response rates by VOT for the lowest (83Hz) and the highest (120Hz) f0 values for each task by Korean listeners and English control listeners](image)

When we examine the individual listeners separately, however, there is substantial variation. As an illustration of inter-listener variation, Figure 6 shows the responses of two listeners—Listener A (YOB: 1991, gender: F, location: Seoul) and Listener B (YOB: 1958, gender: F, location: Seoul). Listener A’s response distributions are similar to the aggregate pattern shown in Figure 4—the listener relies on both VOT and f0 to distinguish the contrast, and the response patterns are similar across all three tasks. On the other hand, Listener B’s L2 responses are similar to those of English control listeners and her L2L1 responses closely mirror her L2, rather than L1 responses. Therefore, for participants like Listener B, L1 perceptual strategies cannot provide a satisfactory account for cross-language perception patterns.
To examine the individual variation more systematically, we test how L2L1 perception and its relationship with L1 or L2 perception differs as a function of L2 proficiency. We estimated the level of participants’ L2 knowledge based on the similarity between their L2 perception and the control English listeners’ perception. This proficiency measure was calculated by first building a logistic regression model for the English control group. The model included the choice between aspirated and non-aspirated as the response variable and VOT, f0, and their interaction as predictors. The code is provided in (3). For this and all analyses below, VOT and f0 were centered by z-score transformation. Analyses were conducted in R (R Development Core Team 2016), and full outputs from all models are provided in the Appendix.

(3) English listeners’ English perception model
  glm(Asp.choice ~ VOT * F0, data = English.Control, family = “binomial”)
stop category: the category boundary is more vertical, indicating a greater reliance on VOT and lesser reliance on f0, than the low-proficiency listeners.

Figure 7: L2L1 mapping by L2 perception proficiency

We can quantify how closely individuals’ L2L1 responses (aspirated vs. non-aspirated) match the English control model’s prediction (voiceless vs. voiced), in the same way we calculated the match between the English control and Korean listeners’ L2 above. Figure 8 plots individuals by their L2 proficiency (x-axis) and the similarity between L2L1 response and English control response (y-axis), and we find a significant correlation (Pearson’s product-moment correlation: $r = 0.729$, $t = 9.645$, df = 82, $p < 0.001$). This analysis confirms our hypothesis that the high proficiency listeners’ cross-language mapping is more consistent with (i.e., more frequently matches) the corresponding L2 category than is the case for the low proficiency listeners.

Figure 8: Scatterplot of the correlation between L2 perception proficiency (x-axis) and the similarity of L2L1 mapping to English control (y-axis). The shading marks 95% confidence intervals.

Note that this correlation could have come about by way of English proficiency affecting the individuals’ L1 perception (see footnote 5) rather than individual’s L2 perception directly affecting the L2L1 mapping. Indeed, there is a significant correlation between individuals’ English proficiency and how closely their L1 perception matches the English categories (Pearson’s product-moment correlation: $r = 0.265$, $t = 2.89$, df = 82, $p = 0.015$). In other words, Korean listeners’ L1 perception was influenced by their knowledge of English (Tice & Woodley, 2012; Schmid, 2013; Ahn et al., 2017). However, L1 perception congruence to English categories accounts for only a fraction of L2L1 mapping variation in comparison to L2 proficiency. A linear regression model that predicts the L2L1 mapping’s similarity to English categories by the same individual’s L1 perception similarity to English categories, the adjusted $R^2$ was only 0.066 while the adjusted $R^2$ was 0.528 for a model that predicts the
L2L1 mapping by the L2 perception similarity to English categories (i.e., English proficiency). So, we can estimate that the bulk of L2L1 variation stems directly from variation in individual listeners’ L2 perception rather than via its influence on L1 perception.

We now probe the consistency of L2L1 mapping further by examining listeners’ L2L1 responses in terms of all three Korean response choices. First, we categorized each English-based stimulus as either “voiced” or “voiceless,” based on the majority response by the English control group. We then calculated each participant’s proportion of aspirated, lenis, and fortis responses for the “voiced” and “voiceless” stops separately. Figure 9 shows these response proportions as a function of participants’ English proficiency.

![Figure 9](image-url)

Figure 9: A scatterplot of by-participant L2L1 response proportions (circle: Aspirated, triangle: Lenis, cross: Fortis) for stimuli perceived by the English control group as English voiced (left panel) and voiceless stops (right panel). The x-axis represents the participants’ L2 perception proficiency and the vertical lines indicate the median proficiency. Curved lines show the smooths of the speaker response proportions for each Korean category, and shadings mark 95% confidence intervals.

We tested whether there was a proficiency-based change in responses using mixed-effects logistic regression models predicting each response choice (e.g. lenis vs. “other”) from participants’ proficiency (centered by subtracting the mean proficiency) with participant as a random effect. The codes are provided in (4).

(4) L2L1 perception choice by English proficiency for English voiced and voiceless stops

\[
\text{glm} \text{mer (Asp.choice } \sim \text{ PROFICIENCY.CONT } + (1|\text{sub}), \text{ data } = \text{data.L2L1.voiced, family } = \text{“binomial”})
\]

\[
\text{glm} \text{mer (Len.choice } \sim \text{ PROFICIENCY.CONT } + (1|\text{sub}), \text{ data } = \text{data.L2L1.voiced, family } = \text{“binomial”})
\]

\[
\text{glm} \text{mer (For.choice } \sim \text{ PROFICIENCY.CONT } + (1|\text{sub}), \text{ data } = \text{data.L2L1.voiced, family } = \text{“binomial”})
\]

\[
\text{glm} \text{mer (Asp.choice } \sim \text{ PROFICIENCY.CONT } + (1|\text{sub}), \text{ data } = \text{data.L2L1.voiceless, family } = \text{“binomial”})
\]

\[
\text{glm} \text{mer (Len.choice } \sim \text{ PROFICIENCY.CONT } + (1|\text{sub}), \text{ data } = \text{data.L2L1.voiceless, family } = \text{“binomial”})
\]

\[
\text{glm} \text{mer (For.choice } \sim \text{ PROFICIENCY.CONT } + (1|\text{sub}), \text{ data } = \text{data.L2L1.voiceless, family } = \text{“binomial”})
\]

For voiceless stops, the higher the L2 proficiency, the more consistently English voiceless stops are mapped to aspirated stops of Korean (\(β=-4.108, z=8.126, p<0.001\)) while both lenis and fortis mappings decline with increased proficiency (lenis: \(β=-3.955, z=-7.652, p<0.001\); fortis: \(β=-2.432, z=-2.372, p=0.018\)). For the voiced stops, while the aspirated responses are minimal, with no effect of proficiency (\(β=-1.802, z=-1.197, p=0.231\)), the lenis responses increase (\(β=2.030, z=2.448, p=0.014\)) and the fortis responses decrease with increased proficiency, although not reaching statistical significance (\(β=-1.685, z=-1.518, p=0.129\)). In other words, not only are voiceless stops converging to aspirated stops in high proficiency listeners’ mapping, voiced stops are also converging to lenis stops, favoring a one-to-one mapping between English and Korean categories rather than one-to-many mapping. Interestingly, the convergence of English voiced stops to lenis stops increases
the consistency of phonemic-level mapping when considering the positional variation in the phonetic realization of the phonological categories: while in word-initial position, English voiced stops are ambiguous between Korean lenis and fortis stops, in word-medial position where English lenis stops are allophonically voiced, lenis stops are clear phonetic matches to English voiced stops. This pattern aligns with the diachronic change in voiced stop adaptation. Previous studies on English loanwords in Korean found that the fortis adaptation of English voiced stops is in decline in established loans (Kang 2008; Oh 2017) and the lenis adaptation is emerging as the dominant adaptation, despite the phonetic ambiguity (Kim 2016; Oh 2017; Park and de Jong 2008; Schmidt 1996).

Given the sound change in progress in Seoul (see footnote 6), a question may arise as to how the speaker age interacts with the findings of the current study. As age and L2 proficiency are negatively correlated in our participants—younger participants are more proficient in English perception—we considered the possibility that the observed L2 proficiency effect on cross-language mapping is an epiphenomenon of the age-based L1 perception transferred to L2L1 perception. We think this is not the case for two reasons. First, the proficiency effect remains significant even when Age is added as a covariate to the statistical model. Second, although our participants do show an age-based difference in cue weighting in the L1 task, consistent with the sound change in progress (with younger listeners showing more use of f0 and less use of VOT than older listeners), this age-based difference does not hold for L2L1 perception, in which younger listeners show more sensitivity to both dimensions than older listeners. The proficiency effect on L2L1 mapping therefore cannot be attributed to the sound change-induced age variation.

Now we turn to the second hypothesis posed in (2), namely, that high proficiency listeners should have greater inhibition of their L1 perception, and instead rely more on the perceived L2 category in their L2L1 mapping, compared to low proficiency listeners. For this comparison, we built mixed-effects logistic regression models of L1 and L2 perception and measured how accurately these models predict individuals’ L2L1 mapping.

The models included the choice between aspirated/voiceless and non-aspirated/voiced as the response variable and VOT, f0, and their interaction as fixed-effects predictors. Crucially, the model included by-subject adjustment to the intercept and slopes as random effects to capture the individual listener variation. The codes are provided in (5); the glmer function of the lme4 package (Bates et al. 2017) was used for analysis.

(5) L1 and L2 perception models by Korean listeners

\[
glmer (Asp.choice \sim VOT * F0 + (VOT*F0|sub), \text{data = data.L1, family = “binomial”) } \rightarrow \text{model.L1}
glmer (Asp.choice \sim VOT * F0 + (VOT*F0|sub), \text{data = data.L2, family = “binomial”) } \rightarrow \text{model.L2}
\]

We then calculated the accuracy of each model predicting L2L1 response using the predict() function in the same way explained above for L2 perception proficiency. This calculation produces two accuracy values for each individual, one for the L1 model predicting L2L1 and another for the L2 model predicting L2L1. For example, from Figure 6, Listener A’s L1 and L2 model accuracy values are 85.0% and 82.5%, respectively. These similar values correctly reflect the observation that this listener’s responses do not differ substantially across the three tasks. For listener B, on the other hand, the two models produced very different accuracy values (L1: 43.8%, L2: 88.8%) and L2L1 responses are very similar to L2 but quite dissimilar from L1.

Now Figure 10 plots all 84 listeners’ L1 and L2 model accuracies. The diagonal equality line marks where L1 and L2 models predict L2L1 equally accurately. The fact that most listeners fall close to this line indicates that their perception is similar across all three tasks like listener A in Figure 6. However, where L1 and L2 model accuracies do diverge, the L2 model tends to be more accurate than L1, as shown by the fact that there are more listeners above the equality line (n=50) than below (n=29; for the remaining 5 participants, L1 and L2 performance were equally predictive). This difference is greater for higher proficiency listeners, in line with the hypothesis that more proficient L2 listeners will give relatively more weight to their L2 perception in L2L1 mapping than less proficient L2 listeners. The L2 model outweighs the L1 model for 66% (29 out of 44) of high L2 proficiency listeners (circles) but this is the case for only 42.5% (for 17 out of 40) of low L2 proficiency listeners (triangles).
Based on the exploratory analysis above, it appears that for listeners with low English proficiency, L1 and L2 models are equally accurate in predicting L2L1 perceptual patterns, which is perhaps unsurprising given that the L1 and L2 patterns are relatively similar. On the other hand, higher-proficiency listeners show more distinct L1 vs. L2 patterns, and it appears that L2L1 perception more closely reflects L2 than L1 perception. We tested this using a mixed-effects logistic regression model with a response variable of CORRECT (Correct vs. Incorrect), i.e., whether the model predictions matched the listeners’ actual L2L1 responses for each trial. The first predictor variable was MODEL (L1 vs. L2), i.e., the L1 or L2 task responses. The second predictor variable was L2 perception PROFICIENCY (high vs. low, defined by median split as described above). Proficiency is treated as a binary factor in the statistical analysis to facilitate the post-hoc test of MODEL effects for higher vs. lower proficiency groups. A model with proficiency as a continuous variable produces comparable results. We also included an interaction between MODEL and PROFICIENCY as we expect that the proficiency will affect the relative accuracy of L1 and L2 models differently. We also consider the possibility that the experimental order may be responsible for the asymmetrical influence of L2 vs. L1 primacy shown in some of the listeners. Recall that in our study, the participants completed all three tasks in one visit and the experimental order was counterbalanced across participants: L2L1 was always the last task but the order of L1 and L2 was varied, with half of the participants completing L1 first then L2 (L1-L2-L2L1) and the other half completing L2 first and then L1 (L2-L1-L2L1). So, we can imagine that other things being equal, L2L1 may be more affected by the task immediately preceding L2L1 than the earlier task. To control for the order effect, we included ORDER (L1 first vs. L2 first) and its interaction with MODEL. All three predictors are coded as sum contrasts (MODEL: L1 = -0.5, L2 = 0.5; PROFICIENCY: low = -0.5, high = 0.5; ORDER: L1 first = -0.5, L2 first = 0.5). We also included by-subject adjustment to the intercept and slope (for MODEL) as random effects. The model is given in (6) and the model output is summarized in Table 3. Figure 11 provides a visual summary by plotting by-participant average accuracy of L1 and L2 model in predicting L2L1 responses against the participants’ English perceptual proficiency.

(6) Model of L2L1 prediction accuracy

\[
glmer(CORRECT \sim MODEL \times PROFICIENCY + MODEL \times ORDER + (MODEL|sub), data = Accuracy)
\]

|                  | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------|----------|------------|---------|----------|
| (Intercept)      | 1.167    | 0.043      | 27.388  | < 0.001  *** |
Table 3: Summary of fixed-effect coefficients in the L2L1 prediction accuracy model

<table>
<thead>
<tr>
<th></th>
<th>Model (L2 vs. L1)</th>
<th>Proficiency (High vs. Low)</th>
<th>Order (L2 first vs. L1 first)</th>
<th>Model * Proficiency</th>
<th>Model * Order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.161</td>
<td>0.067</td>
<td>2.393</td>
<td>0.017 *</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.232</td>
<td>0.085</td>
<td>2.724</td>
<td>0.006 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.003</td>
<td>0.085</td>
<td>-0.033</td>
<td>0.973</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.275</td>
<td>0.135</td>
<td>2.044</td>
<td>0.041 *</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.066</td>
<td>0.134</td>
<td>-0.487</td>
<td>0.626</td>
<td></td>
</tr>
</tbody>
</table>

The statistical model found significant main effects of MODEL (β=0.167, z = 2.393, p = 0.017) and PROFICIENCY (β=0.232, z = 2.724, p = 0.006). The interaction of MODEL and PROFICIENCY is also significant (β=0.275, z = 2.044, p=0.041), indicating the MODEL effect is modulated by the proficiency. The experimental ORDER did not have any significant main effect or interaction with MODEL (main effect: β=-0.003, z = -0.033, p = 0.973; interaction: β=-0.066, z = -0.487, p = 0.6263). Figure 11 shows that L2 matches L2L1 more accurately than L1 for high L2 proficiency listeners, while there is no clear difference between L1 and L2 accuracy for low proficiency listeners. A post-hoc Wald Chi-square test with Bonferroni correction using the `testInteractions()` function of the `phia` package (De Rosario-Martinez et al. 2015) shows that the MODEL effect is significant for high proficiency listeners ($\chi^2 = 10.240, p = 0.003$) but not for low proficiency listeners ($\chi^2 = 0.059, p = 1.000$). From Figure 11, we also observe that this group difference stems from a difference in L2 accuracy in predicting L2L1 responses between the groups—L2 is more accurate for high proficiency than low proficiency listeners—while L1 accuracy remains largely comparable. A post-hoc test also supports this observation: the PROFICIENCY

---

9 An anonymous reviewer wondered if using a predictor of length of residence (LOR) in an English-speaking environment would yield a similar result as we found here, using English proficiency as a predictor. To answer this question, the speakers were grouped into short vs. long LOR groups (less than vs. longer than 6 months) and unsurprisingly, a higher proportion of speakers with a long LOR were high-proficiency speakers (9 high vs. 3 low proficiency) than for speakers with a short LOR (35 high vs. 27 low proficiency). We built a model replacing LOR as a predictor in place of proficiency and found no main effect of LOR or any significant interaction, suggesting that the LOR in and of itself is not a good predictor of changes in the impact of L1 and L2 perception on L2L1 perception.

10 `testInteractions(model.fit, fixed = "PROFICIENCY", pairwise = "MODEL", adjustment = "bonferroni")`
effect is significant for the L2 model ($\chi^2 = 15.140, p < 0.001$) but not for the L1 model ($\chi^2 = 0.18, p > 0.863$).\textsuperscript{11} The results confirm our predictions, namely, that L2L1 perception is affected by L2, as well as L1, and that high proficiency listeners rely on L2 relatively more than L1 compared to low proficiency listeners. Moreover, we also found that this proficiency difference in the relative weighting of L2 vs. L1 was due to the fact that the higher-proficiency listeners used L2 more while the accuracy of L1 model remained comparable across low and high proficiency listeners.\textsuperscript{12} This seemingly contradictory result – convergence toward L2 without divergence from L1 – may be due to L1 perception, as well as L2 and L2L1 perception, becoming more similar to English in high proficiency listeners.

To summarize, we confirmed both of our hypotheses set out in (2). We found that more proficient listeners show more consistent mapping of English voiceless and voiced stops to Korean aspirated and lenis stops, respectively, in L2L1 perception, compared with less proficient listeners. At the same time, more proficient listeners’ Korean aspirated stop responses in L2L1 mapping closely approximate their own L2 perception of English voiceless stops than their own L1 perception of Korean aspirated stops, compared with less proficient listeners. While we found some evidence that more proficient listeners’ L1 perception becomes more congruent to English categories, this change in L1 does not explain the proficiency-conditioned change in L2L1 mapping. Instead, high proficiency listeners’ L2L1 mapping relies on the perceived L2 categories, and these findings are in line with Chang (2015)’s proposal that in experienced L2 users, L1-L2 equivalences follow phonemic rather than acoustic similarity.

5 Conclusion
The current study examined the role of listeners’ L2 phonological knowledge on cross-language perception and probed the relationship between cross-language mapping and L1 and L2 perception within the same individuals. First, we established that L2L1 perception indeed varies as a function of L2 proficiency, i.e., listeners who perceive the English contrast in a more native-like way also show more consistent mapping of English voiceless stops to Korean aspirated stops and English voiced stops to Korean lenis stops in L2L1 perception. We then found evidence that the more proficient in L2 the listener is, the more they rely on their L2 sound categories in cross-language mapping. In other words, our experimental data suggest that L2 knowledge mediates cross-language perception, especially for higher proficiency listeners. The results are significant in that they show how cross-language mapping is constrained by the (perceived) phonological categories of the L2 input and that the phonological structure of L2 can play a crucial role (Paradis and LaCharité 1997), even if the underlying mechanism of loanword adaptation is perceptual in nature.

Another contribution of this study is that it compared L1, L2, and L2L1 perceptions in a controlled acoustic space and directly tested the purported relationship between these three types of mappings within the same individuals.

Our findings have important implications for understanding the mechanism of loanword adaptation. This L2-mediated model of L2L1 mapping is one way to resolve the paradox that loanword adaptation shows sensitivity to perceptual similarity between the source form and the native output but, at the same time, that established loans are more consistent and systematic than naïve filtering by L1-based perception would predict. The findings also add to previous literature highlighting the joint contribution of phonological and phonetic knowledge of the source language in loanword adaptation (Smith 2009, Kang 2010, de Jong and Cho 2012, Chang 2012). The multi-dimensional nature of cross-language perception and its interaction with L2 proficiency also fits in nicely with the observed effect of community-level bilingualism on adaptation patterns (Haugen 1950; Heffernan 2007). The correlation between the level of bilingualism in the community (low vs. high) and the predominant pattern of loan adaptation (phonic vs. phonological) has a parallel with the individual-level connection between

\textsuperscript{11} testInteractions (model.fit, fixed = “MODEL”, pairwise = “PROFICIENCY”, adjustment = “bonferroni”) 
\textsuperscript{12} Comparable results hold with proficiency as a continuous predictor, and no effect of proficiency was found on L1 model accuracy in predicting L2L1 response ($\chi^2 = 0.264, p = 1.000$), while a significant effect of proficiency is found on L2 model accuracy in predicting L2L1 response ($\chi^2 = 13.952, p < 0.001$). testInteractions (model.fit, fixed = “Model”, slope = “Proficiency.cont”, adjustment = “bonferroni”).
proficiency and consistency in adaptation vis-à-vis L2 phonological categories found in the current work. It seems that in cross-language perception, listeners do not suppress their knowledge of L2 categories; furthermore, our findings show that not only do listeners’ perceived L2 categories matter, they override L1 perception in more proficient L2 listeners.

Appendix

a. L1 perception model

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (intercept) | -0.274 | 0.146 | -1.882 | 0.06 . |
| f0 | 1.541 | 0.077 | 19.894 | <0.001 *** |
| VOT | 2.279 | 0.101 | 22.562 | <0.001 *** |
| f0*VOT | 0.266 | 0.046 | 5.781 | <0.001 *** |

b. L2 perception model

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (intercept) | 0.815 | 0.129 | 6.32 | <0.001 *** |
| f0 | 1.501 | 0.079 | 18.954 | <0.001 *** |
| VOT | 1.602 | 0.097 | 16.552 | <0.001 *** |
| f0*VOT | 0.193 | 0.052 | 3.684 | <0.001 *** |

c. L2L1 perception model

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (intercept) | 0.596 | 0.145 | 4.116 | <0.001 *** |
| f0 | 1.279 | 0.070 | 18.320 | <0.001 *** |
| VOT | 1.800 | 0.105 | 17.198 | <0.001 *** |
| f0*VOT | 0.394 | 0.047 | 8.301 | <0.001 *** |

d. English control group model

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (intercept) | 4.795 | 0.311 | 15.438 | <0.001 *** |
| f0 | 0.443 | 0.306 | 1.448 | 0.148 |
| VOT | 4.371 | 0.279 | 15.657 | <0.001 *** |
| f0*VOT | 0.237 | 0.276 | 0.860 | 0.390 |

References


De Rosario-Martinez, Helios, John Fox and R_Core_Team. 2015. phia package, version 0.2-1. Available from https://cran.r-project.org/web/packages/phia/.


Holliday, Jeffrey J. 2016. Second language experience can hinder the discrimination of nonnative phonological contrasts. Phonetica 73, 33–51.


