Draft: September 29, 2017

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

Yoonjung Kang, University of Toronto Scarborough
Jessamyn Schertz, University of Toronto Mississauga

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 L2 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 even 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 phonology 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] would be a fine Japanese word. 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: namely, the long and intense vowel-like release of final nasals in French 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

* Acknowledgements suppressed for anonymous review.

schematically represented by the (blue) dotted arrow in Figure 1.[1] Following (Boersma and Hamann 2009), [] represents phonetic forms while // represent surface, not underlying, phonological forms in this figure.

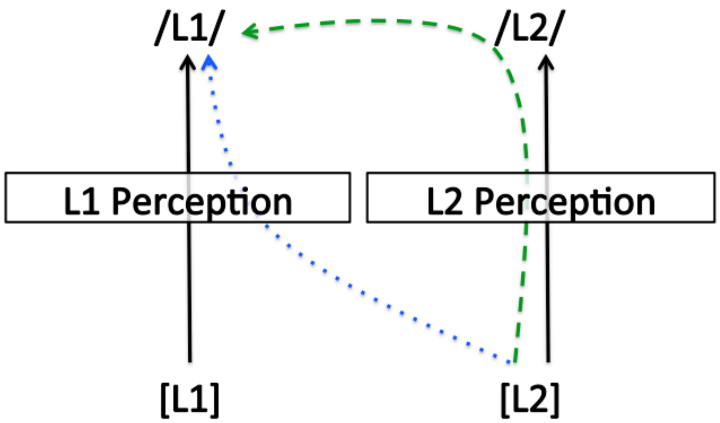

Figure 1: Models of perceptual mapping in loanword adaptation; The blue dotted arrow represents cross-language mapping by a naïve listener without any knowledge of the source language (L2); The green dashed arrow represents a cross-language mapping via listeners' perception/knowledge of source language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

This adaptation-as-naïve-perception model of loanword adaptation is in stark 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, recognizing that the adapters' knowledge may not be native-like and including L2 knowledge as one of various factors that affect the outcome of adaptation (de Jong 2012; Ito 2014; Kang 2010; Kwon 2017; Smith 2006). 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 period of "pre-bilingual period" to a "period of adult bilingualism". Similarly, based on diachronic variation of 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.

In particular, 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) and (de Jong 2012) proposed that while perceptual similarity plays a primary role in shaping adaptation, the variability inherent in cross-language perception is modulated by speakers' knowledge of the phonological categories of the source language, which exerts a regularizing pressure. Specifically, Kang (2010) proposes a constraint named UNIFORMITY which promotes consistent exponence of source language phonological structure leading to convergence toward the most common form among existing adaptation variants. Kang (2010) envisions this regularizing force as a grammatical constraint that overrides perceptually faithful mapping and the regularization itself as a diachronic process that operates on pre-existing loan variants. An alternative possibility is that cross-language perception proper may be affected by listener's source language knowledge. It is this latter possibility that we explore in this study.

Our study builds on the insight from recent findings that cross-language perception is modulated by listeners' knowledge and experience with the input language (Best & Tyler 2007, Bundgaard-Neilson et al. 2011, Nomura and Ishikawa 2016; Kwon 2017). Specifically, Best and Tyler (2007) extended their model of nonnative perception (PAM) to second language (L2) perception (PAM-L2) and proposed that while listeners perceive and

[1] See the figures in Silverman (1992, p. 293), Peperkamp et al. (2008, p. 154), and Boersma and Hamann (2011) for the specific proposals from each model.

process unfamiliar language sounds at the phonetic level, L2 listeners may process L2 sounds at the phonological as well as the phonetic level. For example, using a typical paradigm to test cross-language perception, Bundgaard-Neilson 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* [ho**ms**ɪk] had a target mora of ム [mu]). 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 (cf. 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. However, presence vs. absence of release had no effect on early bilinguals' 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.

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.[2] As a result, adapters with some level of L2 knowledge may first map the L2 signal to L2 phonological structures through their L2 decoder, and these perceived L2 categories are in turn mapped to equivalent L1 categories.[3] This L2-mediated route of perception is schematically represented by the green dashed line in Figure 1.

However, in most previous studies on cross-language mapping, the purported connection between cross-language mapping on the one hand and L1 or L2 perception on the other hand is inferred rather than directly tested or demonstrated. These studies generate predictions about how L2 sounds may be mapped to L1

[2] Yet another possibility why L2 experience affects L2L1 mapping is that L1 perception itself changes due to contact with L2 (Hamman and Li 2016). Under this view, the effect of L2 experience on L2L1 mapping may be via the change in L1 perception proper, and we still predict a congruence between L1 perception and L2L1 mapping and no independent influence of L2 perception on L2L1 perception.

[3] It is not clear from the cross-language perception literature, however, how these /L2/-to-/L1/ phonological equivalences are determined. Best and Tyler (2007) say that L1 and L2 sounds may be identified as "functionally equivalent" and yet not necessarily be "perceived as identical at the phonetic level". For concreteness, we assume that the initial /L2/-/L1/ equivalence itself is grounded in perceptual similarity, with the majority or the best-exemplar non-native perception mapping establishing the equivalence (cf. Kang 2010).

categories based on the comparison of production distribution of L1 and L2 sounds, rather than directly comparing the listeners' perception in L1 and L2 proper. The current study directly tests the assumption of congruence between L1 perception and L2L1 perception that forms the basis of 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 the 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 higher 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.[4] The distribution of Korean and English stops in the VOT * 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.

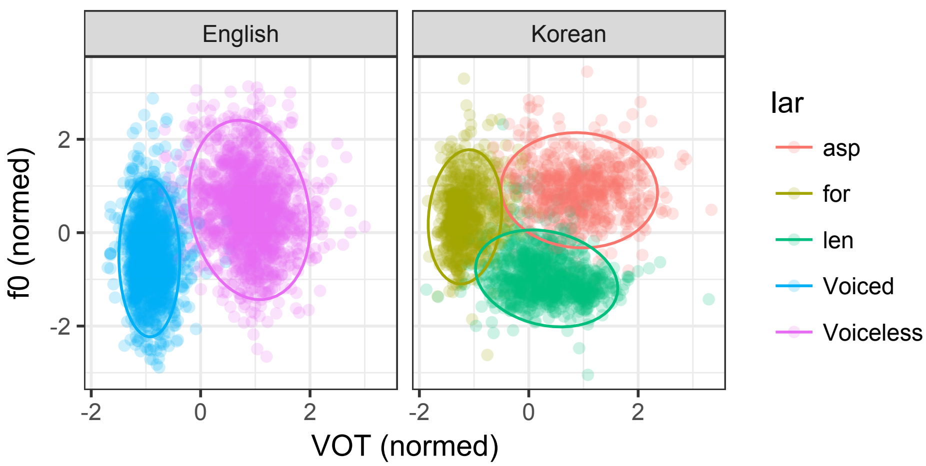

Figure 2: Distribution of English and Korean word-initial stops in the VOT-by-f0 acoustic space

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 aspirated 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; 오미라 2017).

(1)	English	Korean	
	pen	pʰen	(English voiceless -> Korean aspirated)
	beer	piʌ	(English voiced -> Korean lenis)
	bus	p'ʌs'ɨ	(English voiced -> Korean fortis)

In this paper, we focus on the correspondence between English voiceless and Korean aspirated stops. The acoustic similarity between these two classes of sounds in Korean and English (phonetically voiceless, aspirated stops) makes them natural candidates for phonological correspondence. At the same time, as shown in Figure 2,

[4] Seoul Korean (and many other dialects of Korean) is undergoing a sound change in which younger speakers produce a more reduced VOT contrast between lenis and aspirated stops, instead using f0 as the primary cue for the contrast (Kang 2014; Silva 2006).

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), 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, we expect to find variability in perception of English L2 voiceless stops by L1 Korean listeners, with more proficient listeners showing more native-like perception (Kong & Yoon, 2013). This variability allows us to test our hypotheses about how L2 perceptual proficiency (as estimated by 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 grammar 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 English sounds and chose 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

Hypothesis	**Expectations for the current work**
During cross-language perception, more proficient L2 listeners have a **more consistent mapping** between cross-language phonological correspondents.	High-proficiency listeners' perception of Korean aspirated stops on the L2L1 task will more closely approximate native English voiceless stop perception.
More proficient listeners have a **higher relative reliance on L2** (more reliance on L2 and/or more inhibition of L1) in cross-language perception.	More proficient listeners' L2L1 mapping results will more closely approximate their own L2 (vs. L1) perception.

This L2-mediated model of L2L1 mapping can provide a solution to the paradox noted by (Kang 2010) and (de Jong 2012), namely, that loanword adaptation shows sensitivity to perceptual similarity between the source form and the native output but at the same time established loans are more consistent and systematic than naïve filtering by L1-based perception 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 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-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.

Listener group	Task	Stimuli	Response choices
Korean	L1: Korean perception	Korean	Korean categories: (바/빠/파)
	L2: English perception	English	English categories: (ba/pa)
	L2L1: Cross-language perception	English	Korean categories: (바/빠/파)
English	English Control	English	English categories: (ba/pa)

Table 1: Perception tasks by listener group

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 and 21 in Toronto, Canada.[5] All but two of the Seoul participants were "functional monolinguals" who never resided in an English speaking country for more than 6 months (cf. Best and Tyler 2007). Toronto participants varied in their length of residence—11 of the participants were short-term visitors who recently arrived in Toronto to study English, 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). After controlling for L2 perception proficiency, Seoul and Toronto groups behaved similarly and hence we pool the data from the two groups below, unless noted otherwise. Four participants were excluded from analysis; one participant chose a single response for all stimuli, two participants showed random responses, and we exclude one additional listener as her L2 responses were the opposite of the expected direction (high VOT stimuli elicited voiceless responses). Demographic information for the remaining 84 participants is provided in Table 2. 10 English listeners with no knowledge of Korean, recruited in Toronto, participated as a control group.

City	Decade of Birth							Total
	30s	40s	50s	60s	70s	80s	90s	
Seoul	2M, 2F	5M, 5F	4M, 7F	4M, 9F	4M, 1F	1M, 2F	10M, 7F	63
Toronto	-	-	-	-	-	4M	7M, 10F	21

Table 2: Korean participant distribution by recruitment location, decade of birth, and gender.

2.3 Stimuli

Parallel nonsense words in Korean [paru]/[pʰaru]/[p'aru] (바루/파루/빠루) and English [paru]/[baru] were recorded by one male native speaker of each language. Nonsense words with intervocalic [r] were chosen to provide a stimulus-internal cue to the language; between vowels, Korean [r] is realized as a tap while English [r] is an alveolar approximant. Creation of stimuli was identical for Korean and English. We first created baseline tokens for manipulation by splicing aspiration from an aspirated (Korean) or voiceless (English) token onto two vowels varying in voice quality.[6] These baseline tokens were equalized for duration and intensity in each language, then manipulated using the PSOLA algorithm in Praat 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.

[5] Seoul data were collected as part of a larger study on sound change in Korean. Special thanks to XXX for their assistance with data collection.

[6] Vowel quality, as usually 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 (e.g., one fortis and one aspirated vowel for Korean) in order to control for this. As there were no substantial differences in response patterns based on vowel, it is not considered as a factor in the current analysis.

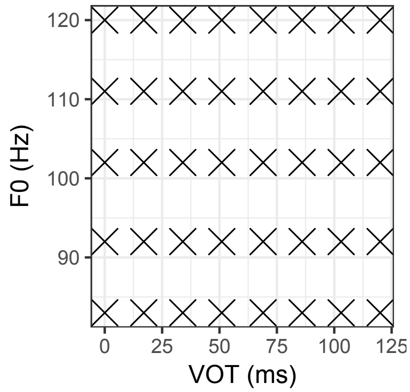

Figure 3: Distribution of stimuli in the VOT-by-f0 acoustic space

2.4 Procedure

Listeners heard the nonce-word stimuli (‘paru’) and responded with the best-fit Korean (p/p’/p^h) 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. L1 and L2 tasks were completed before the L2L1 task, and the order of L1 and L2 tasks varied 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 separate 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, as indicated by the steepness of the curves, 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 is strongly influenced by their L1 perception. This is as expected given that the majority of the speakers in the study are functional monolinguals with limited experience with English.

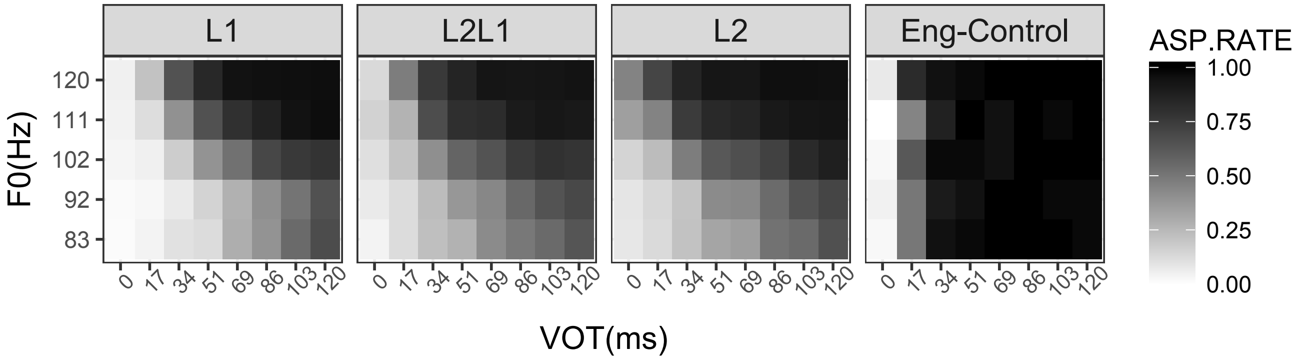

Figure 4: Average aspirated/voiceless response rates for each task by Korean listeners and English control listeners

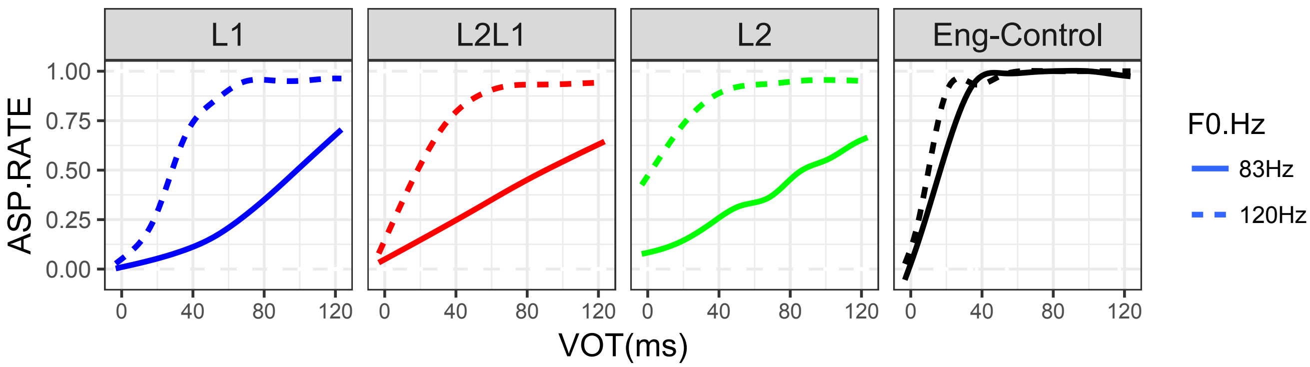

Figure 1: Average aspirated response rates by VOT conditions for the lowest (83Hz) and the highest (120Hz) f0 values for each task by Korean listeners and English control listeners

When we examine the individual listeners separately, however, there is substantial variation. As an illustration of inter-listener variation, Figure 6 shows the response distribution by 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 contrast relies on both VOT and f0 and the responses are distributed similarly 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.

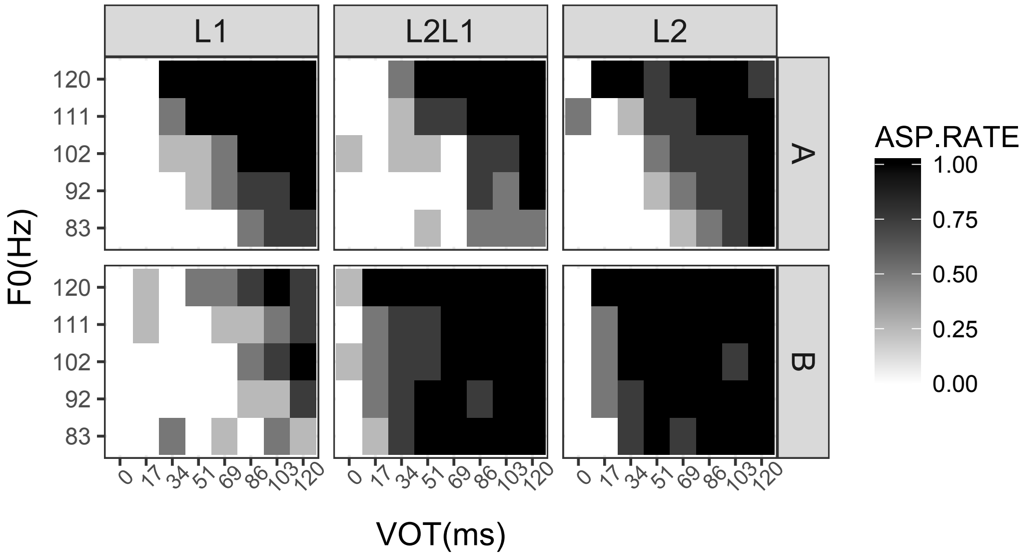

Figure 6: Aspirated/voiceless response rates for each task by two listeners

To examine the individual variation more systematically, we test how L2L1 perception and its relationship with L1 or L2 perception changes 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 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 codes are 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 output from all models are provided in the Appendix.

(3) English listeners' English perception model
glm(Asp.choice ~ VOT * F0, data = English.Control, family = "binomial")

We then compared this English control model's prediction to Korean listeners' L2 (=English) perception, using the *predict()* function in the *stats* package. The model predictions produce a value between 0 and 1, an estimated likelihood of Aspirated/Voiceless responses for each cell of the acoustic space for individual listeners. This proportion is converted to a binary choice of aspirated (>=0.5) or non-aspirated (<0.5) responses. If a listener's actual choice matches the model prediction, it is categorized as accurate and if not, as inaccurate. We then calculate the accuracy across the entire set of stimuli for each listener. We assume that the more closely the Korean listener's perception match that of English control listeners, the more perceptually proficient in English the Korean listener is. The L2 proficiency, as currently defined, showed a substantial amount of variability, ranging from 36% to 94%.

With this proficiency measure in hand, we can now test the hypotheses laid out in (2). Specifically we ask the following questions; (a) are high proficiency listeners' L2L1 mapping more consistent with respect to English categories than low proficiency listeners'?; (b) does L2 exert a relatively greater influence than L1 on the L2L1 mapping of high-proficiency compared to low-proficiency listeners?

First, we test the consistency of the cross-language phonological mapping in L2L1 perception. Figure 8 visualizes the L2L1 responses by high and low proficiency listeners, divided at the proficiency median, and we find that high proficiency listeners show a more English-like category boundary (i.e., more vertical) than low proficiency listeners. In other words, when confronted with a sound that falls into the (native) English voiceless stop category, higher-proficiency listeners are more likely to map it to a Korean aspirated stop, even if it is more phonetically similar to a different Korean category (e.g. lenis). 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 listeners' L2 above. Figure 9 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: $r^2 = 0.532$, t = 9.645, df = 82, p <0.001). This analysis confirms our hypothesis that the high L2 proficiency listeners' cross-language mapping is more consistent with (i.e., better matches) the corresponding L2 category.

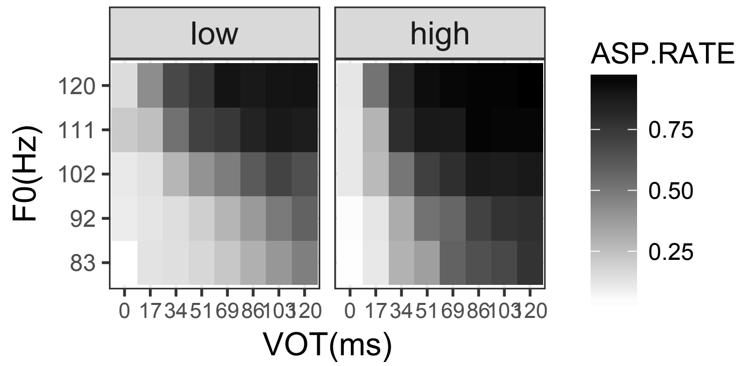

Figure 8: L2L1 mapping by L2 perception proficiency

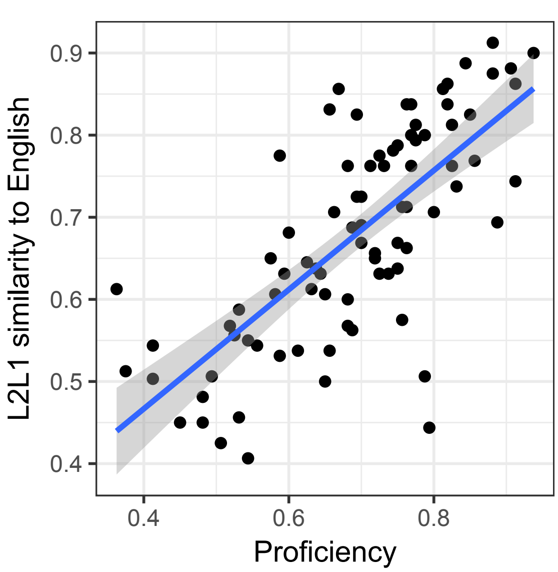

Figure 9: Correlation between L2 proficiency and L2L1 mapping similarity to English control

Now we turn to the second hypothesis posed in (2), namely, if high proficiency listeners are able to suppress their L1 perception and rely on the perceived L2 category more 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 (4); the *glmer* function of the *lme4* package (Bates et al. 2017) was used for analysis.

(4) L1 and L2 perception models by Korean listeners
glmer (Asp.choice ~ VOT * F0 + (VOT*F0|sub), data = data.L1, family = "binomial") → model.L1
glmer (Asp.choice ~ VOT * F0 + (VOT*F0|sub), data = data.L2, family = "binomial") → 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 L1 model predicting L2L1 and another for 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. Most listeners cluster around the upper-right quadrant of the graph falling close to the equality line (L2 model accuracy = L1 model accuracy) meaning that their perception is similar across all three tasks like listener A. This is because for most of our participants, L1 has a pervasive influence on both L2 and L2L1 perceptions. Among those with more dissimilar accuracies across the two models, more are found above the equality line than below, meaning that their L2L1 perception is more similar to L2 perception than L1 perception, like listener B. On the other hand, we don't find as many participants below and far away from the equality line, i.e., participants whose L1 and L2 perception are distinct while L2L1 perception is more similar to L1 perception than L2 perception.

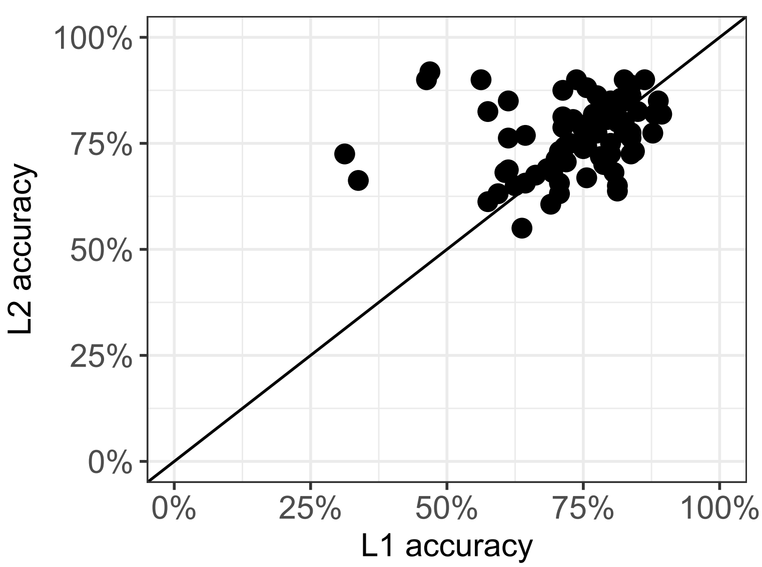

Figure 10: Accuracy of L2L1 perception prediction by L1 and L2 perception models

We hypothesized that more proficient L2 listeners will give relatively more weight to their L2 perception in L2L1 mapping than L1 listeners, and the more knowledge and experience an individual has with L2, the more similar their L2L1 perception will be to their L2 perception. Based on the exploratory analysis above, we predict that for listeners with low English exposure/proficiency, L1 and L2 models are comparable in accuracy of predicting L2L1 because their L1 permeates all three perception tasks, making them all similar. On the other hand, for listeners with more English exposure/proficiency, their L1 and L2 perception will be more distinct and L2 will more closely match L2L1 than L1.

To test these predictions, we built a mixed-effect logistic regression model. The response variable was 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, i.e, the L1 or L2 perception model predicting the L2L1 responses. The second predictor variable was individual listeners' L2 perception PROFICIENCY. Participants were divided into low vs. high proficiency groups divided at the proficiency median. 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[7]. 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 *testInteractions*() function of the *phia* package (De Rosario-Martinez et al. 2015) is used for post-hoc tests. The model is given in (5) and the model output is summarized in Table 3. Figure 9 provides a visual summary by plotting by-participant average accuracy of L1 and L2 model for L2L1 responses against the participants' English perceptual proficiency.

(5) Model of L2L1 prediction accuracy
glmer(CORRECT ~ MODEL* PROFICIENCY + MODEL * ORDER + (MODEL|sub), data = Accuracy)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.255	0.059	21.377	< 0.001 ***
Model (L2 vs. L1)	0.322	0.090	3.581	< 0.001 ***
Proficiency (High vs. Low)	0.182	0.087	2.088	0.037 *
Order (L2 first vs. L1 first)	-0.013	0.087	-0.147	0.883
Model * Proficiency	0.342	0.133	2.572	0.010 *
Model * Order	-0.080	0.132	-0.606	0.545

Table 3: Summary of fixed-effect coefficients in L2L1 prediction accuracy model

The statistical model found significant main effects of MODEL ($z = 3.581$, $p = 0.010$) and PROFICIENCY ($z = 2.088$, $p = 0.037$). The interaction of MODEL and PROFICIENCY is also significant ($z = 2.572$, $p=0.010$), indicating the the MODEL effect is modulated by the proficiency. The experimental ORDER did not have any significant main effect or interaction with MODEL ($p > 0.1$). 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 test with Bonferroni correction shows that MODEL effect is significant for high proficiency listeners ($p < 0.001$) but not for low proficiency listeners ($p > 0.1$). From Figure 11, we also observe that this group difference stems from difference in L2 accuracy 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 that PROFICIENCY effect is significant for the L2 model ($p < 0.001$) but not for L1 model ($p > 0.1$). The results confirm our predictions, namely, L2L1 perception is affected by L2, as well as L1, and the 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 higher-proficiency listeners used L2 more, not that they used L1 less.

[7] 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 others things being equal, L2L1 is more affected by the task immediately preceding L2L1 than the earlier task.

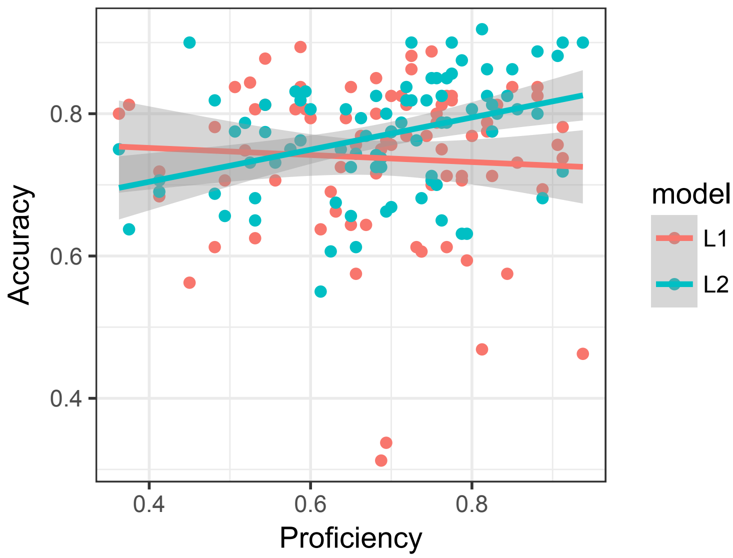

Figure 11: Accuracy of L2L1 prediction by L1 and L2 perception models. Lines show the smooths of the speaker mean accuracies and shadings mark 95% confidence intervals.

4 Conclusion

The current study examined the role of listeners' L2 perception 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 conditioned by individual's 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 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 showing 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 when 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 same individuals.

Our findings have important implications for understanding the mechanism of loanword adaptation. This L2-mediated model of L2L1 mapping provides one way to resolve the paradox noted by (Kang 2010) and (de Jong 2012), namely, that loanword adaptation shows sensitivity to perceptual similarity between the source form and the native output but at the same time established loans are more consistent and systematic than naïve filtering by L1-based perception would predict. 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 level of bilingualism in the community (low vs. high) and the predominant pattern of loan adaptation (phonetic vs. phonological) has a parallel with the connection between proficiency and the consistency in adaptation vis-à-vis L2 phonological categories at the level of individual. It seems that in cross-language perception, listeners do not surpress their knowledge of L2 categories; furthermore, our findings show that not only do listeners' perceived L2 categories matter, they override the 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

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai and Gabor Grothendieck. 2017. lme4 package, version 1.1-13 [Comupter software]. Available from https://cran.r-project.org/web/packages/lme4/.

Best, Catherine T. 1995. A Direct Realist View of Cross-Language Speech Perception. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. Winifred Strange, 171-204. Baltimore: York.

Boersma, Paul and Silke Hamann. 2009. Loanword adaptation as first language phonological perception. In *Loan Phonology*, ed. Andrea Calabrese & W. Leo Wetzels, 11-58. Amsterdam: John Benhamins.

Bundgaard-Nielsen, Rikke L, Catherine T Best and Michael D Tyler. 2011. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition* 33: 433-461.

de Jong, Kenneth, Mi-Hui Cho. 2012. Loanword phonology and perceptual mapping: comparing two corpora of Korean contact with English. *Language Variation and Change 3*: 88.

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/.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier and Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25: 1568-1578.

Haugen, Einar. (1950). The analysis of linguistic borrowing. *Language* 26: 210-231.

Heffernan, Kevin. (2007). The role of phonemic contrast in the formation of sino-japanese. *Journal of East Asian*

Linguistics 16: 61-86.
Ito, Chiyuki. 2014. Loanword accentuation in Yanbian Korean: a weighted-constraints analysis. *Natural language & linguistic theory* 32: 537-592.
Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20: 219-273.
—. 2008. Tensification of voiced stops in English loanwords in Korean. *Harvard Studies in Korean Linguistics* 12: 179-192.
—. 2010. The emergence of phonological adaptation from phonetic adaptation: English loanwords in Korean. *Phonology* 27: 225-253.
—. 2011. Loanword phonology. In *Companion to Phonology*, ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, 2258-2282. Malden, MA.: Wiley Blackwell.
—. 2014. Voice onset time merger and development of tonal contrast in seoul korean stops: A corpus study. *Journal of Phonetics* 45: 76-90.
Kenstowicz, Michael. 2007. Salience and similarity in loanword adaptation: a case study from Fijian. *Language Sciences* 29: 316-340.
Kim, Hyoju. 2016. Contextual distribution of English loanword word-initial tensification in Korean. *음성 음운 형태론 연구* 22: 245-288.
Kong, Eun Jong & Yoon, Inn Hee. 2013. L2 proficiency effect on the acoustic cue-weighting pattern by Korean L2 learners of English: Production and perception of English stops. *Journal of the Korean Society of Speech Sciences*, 5(4):81–90.
Kwon, Harim. 2017. Language experience, speech perception and loanword adaptation: Variable adaptation of English word-final plosives into Korean. *Journal of Phonetics* 60: 1-19.
LaCharite, Darlene and Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation *Linguistic Inquiry* 34: 223-258
Levy, Erika S. 2009. Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America* 125: 1138-1152.
Nomura, Jun and Keiichi Ishikawa. 2016. Effects of first language processes and representations on second language perception: The case of vowel epenthesis by Japanese speakers. *International Journal of Bilingualism*: 1-19.
Paradis, Carole and Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33: 379-430.
Park, Hanyong and Kenneth J de Jong. 2008. Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics* 36: 704-723.
Peperkamp, Sharon, Inga Vendelin and Kimihiro Nakamura. 2008. On the Perceptual Origin of Loanword Adaptations: Experimental Evidence from Japanese. *Phonology* 25: 129-164.
R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. Version 3.3.2. Vienna, Austria: R Foundation for Statistical Computing.
Schertz, Jessamyn. 2014. The structure and plasticity of phonetic categories across languages and modalities. Ph.D. dissertation, University of Arizona.
Schmidt, Anna Marie. 1996. Cross‐language identification of consonants. Part 1. Korean perception of English. *The Journal of the Acoustical Society of America* 99: 3201-3211.
Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary korean. *Phonology* 23: 287-308.
Silverman, Daniel. 1992. Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology* 9: 289-328.
Smith, Jennifer L. 2006. Loan phonology is not all perception: Evidence from Japanese loan doublets. *Japanese/Korean Linguistics* 14: 63-74.
Steriade, Donca. 2001. The phonology of perceptibility effects: the P-map and its consequences for constraint organization. *Ms., UCLA*.
Yun, Suyeon. 2016. A theory of consonant cluster perception and vowel epenthesis. Ph.D. dissertation, Massachusetts Institute of Technology.
오미라. 2017. 영어 유성 자음의 경음 차용에 대한 한국어 음운제약 효과. *언어학* 25: 121-141.