# Individual differences in phonetic cue use in production and perception of a non-native sound contrast

Jessamyn Schertz[a]*, Taehong Cho[b], Andrew Lotto[c], Natasha Warner[d]

* Corresponding Author: jessamyn.schertz@utoronto.ca, Tel: +1-437-580-6641
[a] Centre for French and Linguistics, University of Toronto Scarborough, 1265 Military Trail, HW413, Toronto, ON M1C 1A4, Canada
[b] Hanyang Phonetics and Psycholinguistics Lab, #104, College of Humanities, Hanyang University, Seoul (133-791), Korea
[c] Department of Speech, Language and Hearing Sciences, The University of Arizona, P.O. Box 210071, Tucson, AZ 85721, USA
[d] Department of Linguistics, The University of Arizona, P.O. Box 210025, Tucson, AZ 85721, USA

**Abstract:**
The current work examines native Korean speakers' perception and production of stop contrasts in their native language (L1, Korean) and second language (L2, English), focusing on three acoustic dimensions that are all used, albeit to different extents, in both languages: voice onset time (VOT), f0 at vowel onset, and closure duration. Participants used all three cues to distinguish the L1 Korean three-way stop distinction in both production and perception. Speakers' productions of the L2 English contrasts were reliably distinguished using both VOT and f0 (even though f0 is only a very weak cue to the English contrast), and, to a lesser extent, closure duration. In contrast to the relative homogeneity of the L2 productions, group patterns on a forced-choice perception task were less clear-cut, due to considerable individual differences in perceptual categorization strategies, with listeners using either primarily VOT duration, primarily f0, or both dimensions equally to distinguish the L2 English contrast. Differences in perception, which were stable across experimental sessions, were not predicted by individual variation in production patterns. This work suggests that reliance on multiple cues in representation of a phonetic contrast can form the basis for distinct individual cue-weighting strategies in phonetic categorization.

**Keywords:**
stop voicing; Korean, phonetic cue weighting; individual differences; L2

**Highlights:**

- We examined Korean speakers' production and perception of L1 Korean and L2 English stop contrasts.
- VOT, f0, and closure duration distinguished the Korean contrast in perception and production.
- Productions of the L2 English voicing contrast were reliably distinguished by both VOT and f0.
- Participants showed different strategies for perception of the L2 English contrast.
- Variability in perceptual patterns was not predicted by individual differences in production.

## 1 Introduction

The concept of the native language as a "filter" through which foreign sounds are perceived and produced is both intuitive and empirically well-established. A large body of experimental work has found that native-like production and perception are particularly elusive when the acoustic dimensions that are most important in defining a given foreign contrast are not used in similar native contrasts. The current work turns to a different situation, focusing on how listeners perceive a foreign contrast defined by acoustic dimensions that *are* used in a similar native contrast, but in very different ways. In particular, we examine native Koreans' production and perception of the stop contrasts in their L1 (Korean) as well as their L2 (English) on three acoustic dimensions: duration of voice onset time (VOT) (a primary cue to the distinction in both languages), fundamental frequency at vowel onset (f0) (a primary cue to the Korean distinction and a secondary cue to the English distinction), and stop closure duration (a secondary cue to the distinction in both languages). Alongside analysis of group patterns of acoustic cue use, we focus on individual variability in both production and perception.

### 1.1 Background

Speech sounds contrast on many acoustic dimensions, and a major challenge for language learners is perceiving and acquiring the specific constellation of acoustic "cues" that define the sound contrasts of the L2. As listeners, L2 learners need to determine which cues are relevant to pay attention to, as well as what relative importance, or "weight," to assign each cue. At the same time, as speakers, they must mirror native speakers' use of the cues in their productions in order to attain native-like pronunciation patterns. The majority of work on L2 sound discrimination has focused on how L2 learners cope with foreign contrasts that rely primarily on cues that are not used in similar native contrasts. In these cases, the challenge of learning novel cue mappings is augmented by the fact that throughout development, listeners lose the ability to make use of "foreign" dimensions as cues to category membership, and it is difficult for adult listeners to learn to redirect their attention to the relevant cue. The notorious difficulty of the English /l/-/ɹ/ distinction for Japanese learners, for example, can be attributed to the fact that the English contrast relies primarily on a difference in third formant (F3) values, whereas Japanese listeners distinguish the categories primarily on the basis of second formant (F2) values (Miyawaki et al. 1975; Iverson et al. 2003). Another well-studied challenge for many L2 English learners is the /i/-/ɪ/ distinction, which native listeners distinguish primarily with spectral differences (Stevens 1959; Hillenbrand et al. 2000). In contrast to native listeners, L2 learners whose native languages do not have spectral distinctions in this region of the vowel space often distinguish this contrast primarily on the basis of duration (e.g. Flege et al. (1997) for Spanish and Mandarin, Kondaurova and Francis (2008) for Russian, Cebrian (2006) for Catalan, Morrison (2008, 2009) for Spanish; see also Bohn and Flege (1990) and Escudero et al. (2009)).

In the examples discussed above, the difficulty for non-native listeners arises from their lack of sensitivity or attention to the cue that is most relevant for native listeners. Less well-studied are cases in which an L1 contrast primarily relies on *more* cues than the corresponding L2 contrast, and cases in which an L1 contrast has a *heavier* reliance on a given dimension. In these situations, instead of "filtering out" dimensions that aren't relevant in native contrasts, the L1 sound system might be expected to "filter in" dimensions that *are* relevant to native contrasts, even though they may be unreliable cues to the L2 contrast and ultimately lead to non-native-like categorization. Examining these sorts of scenarios is therefore important for understanding to what extent listeners learn to ignore inefficient or unreliable cues (arising from L1 bias) during L2 sound category

acquisition. As a case study, we examine how Korean listeners, who use multiple primary cues to make phonemic distinctions in their L1 stop contrast (VOT and f0), use these cues to produce and perceive the English stop voicing contrast, which relies primarily on only one of these dimensions (VOT), as well as how they use closure duration, which is a secondary cue in the stop distinction in both languages. We test the same listeners' use of the three cues across production and perception in both languages in order to be able to examine cross-language patterns and the production-perception link on an individual level, as well as the nature and the extent of individual variability in cue-weighting strategies. After a brief overview of issues of individual variability in L2 phonetic categories and background on the Korean and English stop contrasts, the remainder of the introduction turns to how the parallel cross-language, cross-modal dataset collected in the current work can inform theoretical issues about L1-L2 influence as well as the perception-production interface.

### 1.2 Individual variability in L2 cue weighting

A large body of work has considered linguistic (e.g. influence from the L1) and non-linguistic (e.g. length of exposure to the L2) factors that contribute to differences in L2 cue weighting. For example, one systematic effect of L1 is the fact, discussed above, that L2 learners are generally less successful at directing attention to a dimension not used in their native language. The failure to use spectral cues in vowel contrasts such as English /i/-/ɪ/ is seen in listeners whose L1s do not make use of spectral contrasts for vowels in that region of the acoustic space (e.g. Flege et al. 1997 for Spanish and Mandarin, Cebrian 2006 for Catalan, Kondaurova and Francis 2008 for Russian), whereas L2 learners whose L1 is German, a language that *does* use spectral cues in a similar vowel distinction, show more native-like spectral reliance in production and perception of this contrast (Flege et al. 1997). However, the fact that these L2 listeners instead relied on *durational* cues cannot be directly attributed to L1 transfer, given that the languages in question do not use duration for phonemic vowel distinctions. Therefore, factors other than L1 transfer appear to play a role; this phenomenon has been variously attributed to a focus on duration in L2 pedagogy (Flege et al. 1997), different stages of learning (Escudero and Boersma 2004; Morrison 2008), and overall perceptual salience of duration (Bohn 1995). Furthermore, group effects can mask a large range of individual differences within the same language group. For example, although a main finding from the work of Escudero et al. (2009) was that L1 Spanish-L2 Dutch listeners used durational cues more than spectral cues in distinguishing the Dutch /aː/-/ɑ/ contrast, 14 out of the 38 native Spanish listeners actually used spectral cues more. A further complication is the fact that some listeners actually showed reversals of native-like cue use (e.g. with *longer* duration signalling /ɪ/, vs. /i/, categorization, opposite of the native pattern; see Flege et al. 1997; Escudero and Boersma 2004; Morrison 2008), which is not well-explained by any of the proposed mechanisms (though see Morrison 2009 for a developmental hypothesis).

Similarly, L2 proficiency (along with related factors, such as length of L2 exposure) has been shown to influence L2 cue-weighting strategies. Broad differences in language experience or proficiency on a group level are often linked to differences in perception, and in general, studies that split up early vs. late learners (or more vs. less proficient speakers) show the expected pattern of more native-like cue weighting for more experienced groups. For example, Flege et al. (1997) found that experienced listeners from different language backgrounds both produced and perceived L2 English vowel contrasts with more native-like cue-weighting strategies (see also Flege et al. (1996); Bohn and Flege (1997); Baker and Trofimovich (2005); Kong and Yoon (2013), among many others). However, efforts to find individual correlations between proficiency and native-like perception on an individual level have proven elusive; variation in cue-weighting strategies for L2 vowel con-

trasts by L1 Catalan and Spanish listeners (Cebrian 2006; Morrison 2008) and for the L2 English /ɹ/-/l/ distinction (Hattori and Iverson 2009) by L1 Japanese listeners was not correlated with experience or proficiency. These studies did not have the large numbers of participants required to make strong claims about the effect of proficiency, and direct investigation of the relationship is therefore warranted. Nevertheless, although it seems clear that cue-weighting strategies change during the course of L2 development (see Escudero and Boersma (2004) and Morrison (2009) for specific proposals), proficiency effects do not appear to fully account for the variability found in L2 cue-weighting strategies within similar language groups.

Some work has explored factors that could account for the residual within-group variability not explained by language experience or proficiency. Polka (1991) suggests that different strategies for L1-L2 category mapping (even within a single language group) underlie some of the individual variability found in English listeners' discrimination of the Hindi retroflex-dental stop contrast (on the other hand, Hattori and Iverson (2009) show that differences in L1-L2 mapping do not account for variability in native Japanese identification performance on the English /l/-/ɹ/ contrast). Escudero and Boersma (2004) show that foreign cue-weighting strategies can reflect the specific dialect of the L2 being learned (in this case, Native Spanish listeners' relative use of spectral and temporal cues in the Scottish vs. Southern British English /i/-/ɪ/ contrast). Individual differences in L2 category discrimination have also been linked to differences in sensitivity to *L1* phonetic contrasts: Díaz et al. (2008) found that listeners who performed better at an L2 category discrimination task showed larger mismatch negativity (MMN) responses to both native and non-native sound contrasts. Regardless of the provenance of the differences, some recent studies show that differences in initial categorization strategies can result in different sorts of learning patterns during perceptual training tasks (Chandrasekaran et al. 2010; Wanrooij et al. 2013, Schertz et al. (submitted)), suggesting that these differences in cue weighting strategies can be robust enough to form the basis of different sorts of adaptation strategies.

The emphasis on individual variability in the current work primarily focuses on the question of the stability of individual cue weighting strategies across languages and modalities; in other words, to what extent does an individual show the same use of cues across languages, and across perception and production? If a given individual's cue weighting strategies are consistent across perception and production (or across languages), this suggests that these category definitions are present on some level of representation that is tapped into by both modalities (or languages). We also examine the question of whether individual patterns are stable across time by comparing the performance of a subset of the same participants across experimental sessions on separate days.

### 1.3 Korean and English stop contrasts

Korean and English speakers make use of some of the same acoustic dimensions in defining their native stop contrasts; however, the relative use of these cues is quite different. Korean has a typologically unusual three-way stop contrast: lenis (불 [pul] 'fire'), fortis (뿔 [p*ul] 'horn'), and aspirated (풀 [p$^{h}$ul] 'grass'). The realization of the contrast varies depending on dialect and age (e.g. Silva 2006; Kang 2009), but for young speakers of the Seoul dialect, the population in the current work, the stops contrast in VOT, f0 at vowel onset, closure duration, as shown in Table 1[1]. Although these cues have not all been covaried such that their relative primacy can be determined, previous work has shown that the three-way distinction can be captured in both production (e.g. Lee and Jongman 2012) and perception (e.g. Lee et al. 2013) using a combination of VOT and f0,

[1]See also Cho et al. (2002) and Lee and Jongman (2012) for thorough overviews of previous acoustic and aerodynamic studies; note that voice quality also plays an important role in the Korean stop contrast (Cho et al. 2002; Kang and Guion 2006; Lee and Jongman 2012).

| | Lenis | Fortis | Aspirated | Selected references |
|---|---|---|---|---|
| VOT | Intermediate | Shortest | Longest | Lisker and Abramson 1964; Kim 1994; Cho et al. 2002 |
| | *Although VOT has traditionally been found to distinguish all three stop categories, more recent studies (Silva 2006; Kang and Guion 2006; Lee and Jongman 2012), show the difference in VOT values between lenis and aspirated stops to be increasingly overlapping. | | | |
| f0 | Lowest | Higher | Highest | Kim 1994; Cho et al. 2002; Kang and Guion 2006; Lee and Jongman 2012 |
| Closure duration | Shorter | Longest | Longer | Kim 1994; Cho and Keating 2001 |

Table 1: Acoustic dimensions defining the Korean three-way stop contrast, as documented in previous work.

| | Voiceless | Voiced | Selected references |
|---|---|---|---|
| VOT | Long | Short (or prevoiced) | Lisker and Abramson 1964 |
| f0 | Higher | Lower | Llanos et al. 2013 |
| Closure duration | Longer | Shorter | Green et al. 1998 |

Table 2: Acoustic dimensions defining the word-initial English stop voicing contrast, as documented in previous work.

suggesting that these two cues together provide a sufficient acoustic space for defining the contrast, and that both play important roles in distinguishing the categories (at least in the contemporary dialect of younger speakers from Seoul).

On the other hand, the English phonological stop voicing contrast (e.g. /p/ vs. /b/) is defined primarily by VOT, and to a much smaller extent by other cues, such as f0 at vowel onset and closure duration (Table 2), with these secondary cues being recruited mainly when the primary cue is ambiguous (Francis et al. 2008). In sum, although VOT, f0, and closure duration all play a role in both the Korean and English stop contrasts, the relative use of these cues is very different, raising the question of how native Korean listeners learning English recruit these cues when defining their L2 sound categories. In contrast to having to learn to pay more attention to a cue (as would be the case for an English listener learning Korean), the Korean listeners need to learn to ignore (or "downweight") an acoustic cue that is primary in their L1 but provides less reliable information in their L2, in order to achieve native-like perception and production.

### 1.4 L1 influence on L2 sound categories

Previous work has found that Korean speakers and listeners make use of both VOT and f0 when producing (Kang and Guion 2006; Kong and Yoon 2013) and perceiving (Kim 1994; Kong and Yoon 2013) their L2 English stop contrast, implying that the heavy use of f0 in their native contrast may transfer to their L2 sound categories. However, the group patterns shown in this previous work do not give a detailed picture of how, why, or to what extent this L1 influence is realized. It might be expected that the primacy of both VOT and f0 in the L1 (Korean) contrast has given listeners enhanced sensitivity to these dimensions and their distributional properties, such that their

perception would reflect that of native listeners almost exactly. It is also possible that they would choose a single dimension on which to distinguish the L2 contrast, given that multidimensional sound category learning may be more difficult (Goudbeek et al. 2008, see also Ashby et al. 1999 for visual category learning), even though they use both dimensions in a single native sound contrast. On the other hand, it might be the case that they treat both primary cues to the Korean contrast as primary in the corresponding English sound contrast, even though one of these cues (f0) is only a very weak cue for native English speakers. In the current work, we explore whether Korean listeners' performance is consistent with one of these hypotheses. The focus on individual variability allows us to examine whether the choice is consistent across listeners, or whether, given these multiple options, individual listeners make different choices for which dimension(s) to pay attention to. The use of the same population for perception and production tasks in both languages also allows us to directly compare the extent of native language influence in production as opposed to perception.

L1 influence on L2 can also be examined in light of two major theories of L2 speech sound acquisition: the Speech Learning Model (SLM, Flege 1995, 2007) and the Perceptual Assimilation Model's extension to L2 acquisition (PAM-L2, Best and Tyler 2007). Both of these models rest on the assumption that L2 sounds are assimilated to L1 sound categories whenever possible, and they make predictions about the difficulty of L2 sound discriminability as a function of their phonetic similarity and patterns of assimilation to L1 categories. In the case of native Koreans' perception of the English stop contrast, discriminability should not be an issue. Given that the Korean stop contrast makes use of VOT as a primary cue, English phonemically voiced (/b, d, g/) and voiceless (/p, t, k/) stops will almost certainly be predicted to assimilate to different Korean categories. Both PAM and SLM predict that an L2 contrast with this sort of "two-category assimilation" should be easy to discriminate. Along these lines, Kang and Guion (2006) showed that native Korean speakers who had learned English later in life appeared to have merged the English voiceless and Korean aspirated categories in production; English voiced stops, on the other hand, appeared to be a distinct category from either Korean fortis or lenis stops (similar to fortis stops in terms of VOT, but similar to lenis stops in terms of f0). Similarly, Kong and Yoon (2013) showed that native Korean speakers with higher English proficiency made more effective use of VOT in distinguishing the English stop contrast, and that the high-proficiency group showed more native-English-like performance in perception as well. Further perceptual findings come from Schmidt (1996) and Park and de Jong (2008), who asked native Korean listeners to listen to English syllables and indicate the Korean consonant that best represented the syllable-initial sound. In both studies, the Korean listeners labeled English voiceless stops as being most similar to Korean aspirated stops, while their judgments for English voiced stops were mixed, split between lenis and fortis stops. Schmidt noted that both acoustic differences in the English voiced tokens and individual listener variability in labeling contributed to this split. The structure of the perception experiment in the current work, which manipulates multiple acoustic dimensions systematically, as well the focus on individual variability in the analysis, allows for robust investigation of both of these factors. Furthermore, by looking at the same listeners' perceptual patterns in Korean and English, we can evaluate the extent to which listeners use the same cues to define the English voiceless and Korean aspirated stop categories, as would be predicted if the two categories were assimilated.

### 1.5 L2 perception-production interface

The relative use of acoustic cues can be considered in terms of both perception and production. In production, a given dimension can be a very reliable indicator of category membership, with values perfectly separated by category, or highly unreliable, with overlapping values and only a general overall tendency for values to differ across categories (i.e. more or less "informative," Holt

and Lotto 2006). On the level of perception, listeners can give more or less weight to an acoustic dimension when making categorization decisions. Use of cues in L2 production and perception are reflective of one another to a certain extent, in that broad-level production patterns often mirror perceptual patterns (or vice versa). For example, Flege et al. (1997) showed that on a group level (with participants grouped by native language and level of experience), the informativeness of spectral vs. temporal cues in each group's production of English tense-lax vowel contrasts could to some extent predict cue use in perception of the same contrasts. Links between vowel discrimination and intelligibility have also been found (e.g. Flege et al. 1999; Flege and MacKay 2004; Sebastián-Gallés and Baus 2005). Rochet (1995), Bradlow et al. (1997), and Wang et al. (2003) found that perceptual training alone can result in more native-like productions. However, even broad group differences are often not always consistent across the two modalities. Native Japanese perception of the English /l/-/r/ contrast, as discussed above, relies primarily on F2, as opposed to F3 (Iverson et al. 2003); however, native Japanese *productions* of English /r/-/l/ are actually better differentiated by F3 than F2, although F3 is still less informative, and F2 is more informative, than for native English productions (Lotto et al. 2004). The Korean-English case study allows us to test whether the relative contribution of the different acoustic cues (VOT, f0, and closure duration) are stable across perception and production of the L2 contrast, or if, as in the Japanese example, there is an asymmetry in how L1 influence manifests itself across the two modalities.

Given the expected variability between individuals, an additional question that arises is whether this variability is stable across perception and production in a given participant. Based on proposals that perceptual cue weights emerge from statistical regularities in the input (e.g. Holt and Lotto 2006; Francis et al. 2008; Toscano and McMurray 2010), one might expect that relative informativeness of a dimension in a speaker's productions of a contrast can predict the weight given to that cue in perception of the same contrast. This view also follows from theories of speech processing that posit a strong and/or direct perception-production link (e.g. Liberman and Mattingly 1985; Fowler 1986). However, work looking for perception-production links in terms of individual cue weights has generally failed to find correlations between the two modalities. For example, Bohn and Flege (1997) examined native German speakers' use of spectral vs. durational cues in the production and perception of the English /ɛ/-/æ/ contrast, and found that an individuals' use of durational vs. spectral cues in production did not correspond to their reliance on the two types of cues in perception. With the data collected in the current work, we are able to test the hypothesis that individual variability in relative cue use in production correlates with individual perceptual patterns on a forced-choice perception task (i.e. whether speakers who produce produce large differences in f0 in the English voicing contrast also rely more heavily than average on f0 to distinguish the contrast in a perception task).

### 1.6 Goals of the current study

The phonetically rich Korean 3-way stop contrast provides an opportunity to investigate native language influence on L2 cue weighting strategies from a novel perspective: to what extent do listeners recruit L1 cues to define L2 contrasts when the corresponding L1 contrast makes use of *more* primary cues than are necessary to distinguish the corresponding L2 contrast? In Experiment 1, we examine how the same group of native Korean speakers uses three acoustic dimensions (VOT, f0, and closure duration) when producing stop contrasts in their L1 (Korean) and L2 (English). Experiment 2 turns to the perception of L1 Korean and L2 English stops in terms of the same three acoustic cues: the same group of native Korean listeners participated in two forced-choice identification tasks on separate days (one in Korean, one in English) for stimuli covarying in VOT,

f0, and closure duration. A follow-up study (Experiment 2a) verifies whether individual listeners' perceptual cue weights are stable across experiment sessions on different days.

Since all three acoustic dimensions are used in the L1 Korean stop contrast, the participants may be sensitive to the distributional patterns of the L2 contrast, in which case we would expect to find native-like use of cues in both production and perception of the English contrast (i.e. VOT should be the most reliable indicator of voiced vs. voiceless in speakers' productions and the most heavily-weighted cue in perception, while f0 and closure duration should be weakly informative of category membership in production and should show a small amount of influence on categorization in perception). On the other hand, given the primacy of f0 in the native Korean contrast, we might expect greater use of f0 in the L2 English contrast. Previous work has shown greater-than-native-like use of f0 (e.g. Kim 1994; Kang and Guion 2006; Kong and Yoon 2013); at the same time, work showing differences in L2-L1 mapping of the English to Korean contrast (e.g. Schmidt 1996; Park and de Jong 2008) suggests that there may be considerable variability in perceptual patterns. We build on this past work by examining the extent of variability in production vs. perception, and whether individual differences are stable across sessions. We also explore individual differences present in perceptual cue weighting reflect speaker-specific strategies in how the contrasts are produced, using correlation analyses to determine whether the reliability of a given acoustic dimension in a speaker's productions can predict the same participant's reliance on that cue in the perception task.

## 2 Experiment 1: Production

Experiment 1 examined native Korean speakers' productions of Korean and English word-initial stops on three acoustic dimensions: VOT, f0, and closure duration.

### 2.1 Methods

#### 2.1.1 Stimuli

Visual prompts for speakers' productions consisted of monsyllabic target words embedded in carrier sentences in either English ("I say _.") or Korean ([igʌ _-iejo], 'This is _.'). The target words consisted of 18 minimal pairs (English) and 13 minimal or near-minimal triplets (Korean) differing in initial stop type. The full set of stimuli is given in Tables 10 and 11 in the Appendix.

#### 2.1.2 Participants

21 native Korean speakers (9 female, 12 male, ranging in age from 19-27 years old), all students at Hanyang University, completed the experiment at the Hanyang Phonetics and Psycholinguistics Laboratory in Seoul. All participants had learned English in school (beginning at a mean age of 10.7 years), but none used it on a regular basis. Eight of the participants had studied abroad in an English-speaking country (Australia, Canada, or the United States) for six months to one year; none of the other participants reported having spent over three months in an English-speaking country.

#### 2.1.3 Procedure

The Korean and English sessions were completed on two separate days, and oral and written instructions were provided in the target language on each day. Participants were seated in a sound-proof booth and fitted with a high-quality head-mounted microphone. A Sony PCM-D50

recorder (44.1 kHz sampling rate) was used for the recordings. Sentences were presented in Korean or English orthography on a computer screen using PsychoPy (Peirce 2007) at a steady rate of 3 seconds (Korean) or 2.5 seconds (English) per sentence in order to elicit a relatively stable speaking rate (the rate was slower for Korean because the Korean carrier phrase was longer than the English one). The list of stimuli was presented three times in randomized order.

#### 2.1.4 Acoustic measurements

All measurements were performed with Praat (Boersma and Weenink 2011). Trials in which the speaker misread the word or hesitated before reading the word were excluded from analysis (14 out of 2394 tokens from Korean; 55 out of 2268 from English).

*Closure duration:* Closure duration was manually measured from the end of formant energy in F2 in the spectrogram of the final sound of the carrier phrase preceding the target to the beginning of the target consonant burst in the waveform.

*Positive voice onset time (VOT):* VOT was manually measured from the beginning of the target consonant burst in the waveform to the first zero crossing in the waveform following the onset of periodicity in the following vowel. Prevoicing, which occurred in some English tokens, was also measured; however, only the positive VOT measure is included in the subsequent analyses[2].

*Fundamental frequency (f0):* Fundamental frequency was measured at 5 ms after the onset of periodicity in the vowel (chosen in order to capture f0 as close to the onset of the vowel as possible while still obtaining a reliable pitch track) using the "To Pitch…" function in Praat. In order to minimize errors in the automatic pitch tracker, individual pitch floors and ceilings were set for each speaker based on manual inspection of the signal prior to measurement, and these individual settings were given as input to the Praat algorithm. We also monitored the automatic measurement process visually in order to ensure reliability of the pitch tracker; automatic pitch tracking errors were manually corrected as necessary.

*Prosodic boundaries:* It is well known that prosodic position has an effect on the duration of stop closure (Cho and Keating 2001, 2009), which is often acoustically inseparable from a pause when produced at an Intonational Phrase (IP) boundary. In both the English and Korean carrier phrases, it was possible to produce a prosodic boundary before the target words. Productions were therefore annotated by the first author, indicating whether there was a perceptible IP boundary before the target word. In the Korean productions, speakers were consistent in whether or not they produced an IP boundary throughout the entire production task (10 speakers produced a break while 11 did not); however, in the English production task, some speakers showed variable prosody, sometimes producing a prosodic break and other times not. In the Results section, we examine whether this prosodic variability potentially masks effects of stop closure duration.

[2]In the context of the English carrier sentence ("I say ___"), prevoiced realizations of the target word-initial phonologically voiced stops /b, d, g/ might be expected in some English productions. One of the Korean speakers showed fully voiced closure durations consistently (39 tokens, or 76% of voiced tokens); another produced voicing throughout the closure duration of voiced stops 33% of the time (17 tokens), while an additional four speakers produced fewer than five tokens showing voicing during the closure. Two speakers showed a few tokens of prevoicing (9 and 3 voiced tokens, respectively) that did not continue through the entire stop closure. In total, only 8.4% of the phonologically voiced tokens showed phonetic voicing, and almost half of these came from a single speaker. Given the small number of phonetically voiced tokens, we chose to use only the measure of positive VOT, as opposed to including prevoicing on the same dimension, in order to keep the analyses across languages comparable.

#### 2.1.5 Omitted data

The majority of pitch tracking errors resulted from creaky voice, which occurred in many of the fortis productions (69 fortis tokens, as opposed to 7 lenis and 1 aspirated). This asymmetry is expected given that fortis stops are often produced with creaky voice (e.g. Cho et al. 2002). Some English tokens also exhibited creaky voice (17 phonologically voiced and 20 phonologically voiceless tokens). In all of the English tokens, the creaky voice appeared to be a property of the vowel, especially since the target word was phrase-final and often produced with lower pitch. Tokens with creaky voice were omitted from analyses including f0 as a predictor variable. We also omitted all tokens of one Korean word, 댐 [tem] 'dam,' because most speakers pronounced the stop as a fortis stop, even though it is orthographically a lenis stop (see Kang (2008) for discussion of variation in realization of stops in English loanwords). Finally, we omitted tokens with values of $3SD$ above the mean for each category in closure duration in order to eliminate tokens that may have been hesitations and therefore obscure potential effects of closure duration. Consequently, 34 Korean tokens and 64 English tokens were omitted, leaving a total of 2346 Korean and 2145 English tokens.

#### 2.1.6 Statistical analyses

For graphs and statistical analyses, unless otherwise noted, the values for each dimension were converted to z-scores within each speaker's productions for each language separately (e.g. a subject's values for f0 in all of his/her Korean tokens were centered around zero) in order to remove speaker-specific overall differences in each dimension, such as gender-based variation in f0 range. The Levene test for homoscedasticity for speakers' average values for each acoustic dimension across stop types in both languages showed that variances for the values of VOT were not equal across stop type in either English ($F(1, 40) = 7.67, p = .008$) or Korean ($F(2, 60) = 9.29, p < .001$) (the variances for f0 and closure duration did not differ significantly across stop types in either language, $p > .05$ for all tests). Values for VOT duration were therefore log-transformed to correct for heteroscedasticity. All statistical analyses below use these log-transformed (for VOT) and standardized (for all dimensions) values.

We performed two types of analysis on the production data. First, in order to determine whether speakers as a group produced significantly different values of VOT, f0, and closure duration for the different stop types in each language, we ran within-subjects analyses of variance (ANOVA) on aggregated group data (one data point per subject per stop category). These analyses show how speakers' productions of the contrasts differed with respect to each acoustic dimension. The Korean stops were analyzed (here and throughout this paper) as the three possible two-way contrasts (Fortis vs. Lenis, Aspirated vs. Fortis, and Aspirated vs. Lenis)

We then used Linear Discriminant Analysis (LDA) to model the extent to which the same acoustic cues reliably predicted category membership in each speaker's productions. Since the LDA model is built and tested on the same data, it is descriptive rather than predictive or inferential, and can be thought of as a metric of how well a given dataset can be separated using an optimized linear combination of a given set of dimensions. Using the *lda* function from the *MASS* package in R v. 3.0.2, we calculated the model that best separated each speaker's production data. Each LDA model used speakers' values for VOT, f0, and closure duration to predict the stop type (voiced or voiceless in the English analysis, or aspirated, fortis, or lenis in the Korean analysis). The coefficients from each dimension provide a metric for how much each speaker weights a given dimension in production. Since the ranges for each dimension were standardized for each speaker prior to analysis (using z-scores), the coefficients across dimensions are comparable, with a coefficient of higher magnitude (either positive or negative) corresponding to more reliable separation of categories on

that dimension[3].

### 2.2 Results from Experiment 1: Production

Mean production values and their standard deviations for Korean and English stops on the three acoustic dimensions examined in this study (VOT, f0, and closure duration) are given in Table 3. Standardized values are plotted against each other in each combination of two cues in Figure 1, and individual plots of the L2 English production data are given in the Appendix (Figure 11).

| | **Korean** | | | **English** | |
|---|---|---|---|---|---|
| | **Lenis** | **Fortis** | **Aspirated** | **Voiced** | **Voiceless** |
| VOT (ms) | 76 (21) | 19 (10) | 86 (22) | 27 (14) | 95 (26) |
| f0, male (Hz) | 120 (14) | 155 (24) | 176 (23) | 117 (19) | 141 (21) |
| f0, female (Hz) | 227 (13) | 281 (25) | 306 (25) | 212 (20) | 248 (25) |
| Closure (ms) | 85 (42) | 125 (44) | 103 (42) | 113 (46) | 116 (35) |

Table 3: Summary of means and standard deviations for VOT, f0 (split by gender), and closure duration for the 21 native Korean speakers' productions of English and Korean word-initial stops.

#### 2.2.1 ANOVA results

Results from the group ANOVAs are given in Table 4. Speakers used all three acoustic parameters (VOT duration, f0, and closure duration) in distinguishing productions of each of the possible combinations of two-way stops in Korean, while values for VOT and f0, but not closure duration, revealed statistically significant differences across stop types in English. However, because closure duration measures could be complicated by prosodic boundary effects, we ran a separate ANOVA excluding those participants who produced IP breaks before the target syllable (n=8). This ANOVA, testing the thirteen remaining participants, did show a significant effect of closure duration ($F(1, 12) = 5.12, p < .05$).

| | | **VOT** | | **f0** | | **Closure** | |
|---|---|---|---|---|---|---|---|
| Language | Contrast | $F(1, 20)$ | $p$ | $F(1, 20)$ | $p$ | $F(1, 20)$ | $p$ |
| Korean | Aspirated vs. Lenis | 7.81 | .004 | 5430.26 | $< .001$ | 54.87 | $< .001$ |
| | Fortis vs. Lenis | 4338.23 | $< .001$ | 326.88 | $< .001$ | 247.15 | $< .001$ |
| | Aspirated vs. Fortis | 3842.84 | $< .001$ | 75.48 | $< .001$ | 206.59 | $< .001$ |
| English | Voiced vs. Voiceless | 3107.20 | $< .001$ | 753.49 | $< .001$ | 1.57 | $> .1$ |

Table 4: Results from ANOVAs comparing values of each acoustic dimension across stop types in native Korean speakers' productions of English and Korean stops. Although the effect of stop type on closure duration was not significant in English across all speakers, when speakers who made a prosodic break were omitted from the analysis, the effect was significant (details in text).

[3]The polarity of the coefficients for a given factor are arbitrary, dependent on the ordering of the levels of the independent variable, such that coefficients of 3 and $-3$ indicate the same reliability of separation. In order to be able to compare the relative reliability of coefficients across contrasts, we flipped the polarity of coefficients for some dimensions such that all dimensions would show positive coefficients if they went in the expected direction (based on previous work, as given in Tables 1 and 2).

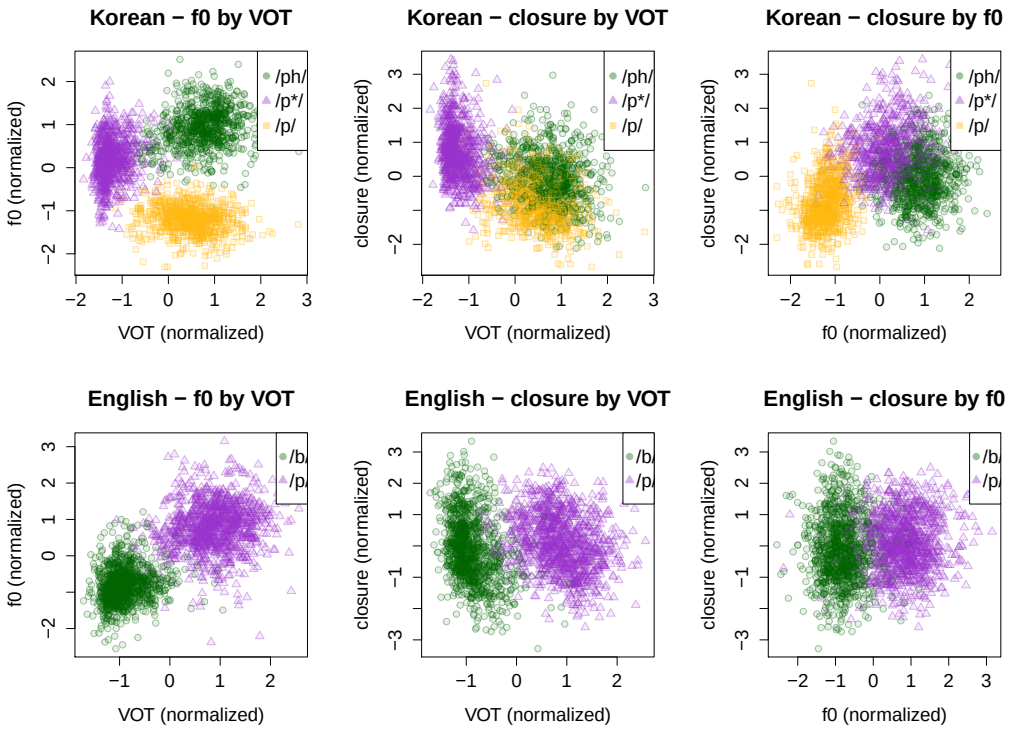



Figure 1: Native Korean speakers' values for all combinations of VOT, f0, and closure duration in the Korean (top) and English (bottom) stop contrasts, standardized (converted to z-scores along each dimension) by speaker. Although values for VOT were log-transformed for the statistical analyses, the non-transformed values are plotted here.

### 2.2.2 LDA results

Graphs showing the distributions of coefficients from the discriminant analysis are given in Figure 2, and a table of all coefficients for all participants is given in Table 12 in the Appendix. Boxplots in Figure 2 show the range of speakers' coefficients for each dimension, with larger coefficients indicating greater separability of the contrast on a given dimension in the expected direction. In general, these coefficients reflect group patterns. The lack of predictability for closure duration in the Korean fortis-lenis contrast, despite a significant effect of closure duration in the ANOVA, is probably due to the fact that closure duration and f0 are highly correlated for that contrast (see Figure 1), so the predictability must be "shared" between the two. For production of the English contrast, although VOT is overall more predictive than f0 in separating the two classes of stops, all participants used both in the expected direction, and the relative use of the two cues differed between individuals: VOT was a more predictive cue for 13 speakers' productions, f0 was more predictive for 7, and the two cues were equally predictive for one speaker[4]. Below we test whether this variability corresponds to variability in cue use in perception.

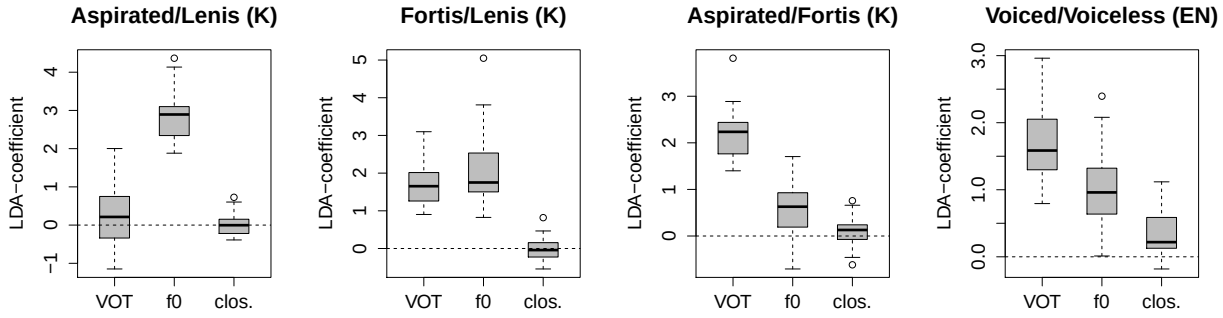


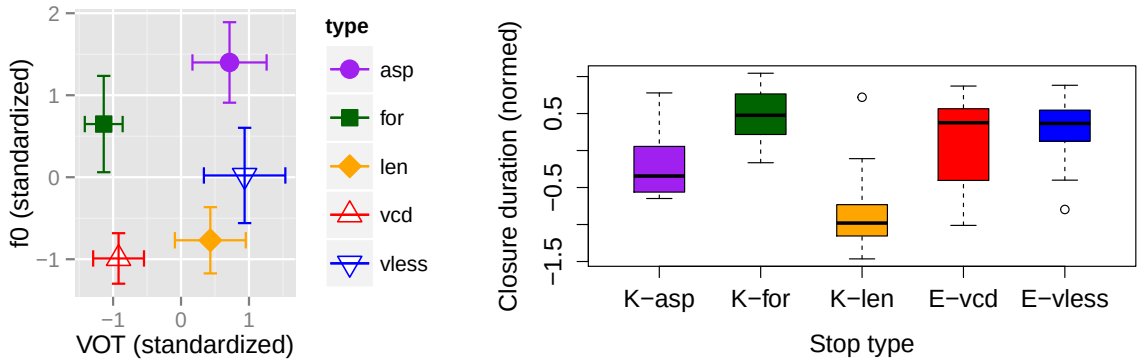

Figure 2: Predictability of stop type based on each cue measured in speakers' productions, as estimated by the distribution of coefficients from individual discriminant analyses for each speaker/contrast. Boxplots show the interquartile range (and any outliers beyond 2 standard deviations from the mean) of individual speakers' coefficients for each dimension. A larger coefficient indicates greater separability of the contrast on a given dimension in the expected direction.

### 2.2.3 Comparison of L1 vs. L2 categories

Figure 3 provides a visual comparison of the acoustic spaces for all English and Korean stops, using values standardized across each speaker's range in both languages combined (i.e. all values of VOT for a given speaker are centered around zero). While there is some overlap on some of the dimensions (e.g. English voiceless and Korean aspirated stops have similar values for VOT),

[4]Under the assumption that more experience in an English-speaking country results in more native-like speech, we might expect those with experience (n=8) to show higher coefficients for VOT and lower values for f0 than those with less experience (n=13). This prediction was not supported for VOT ($t(19) = -1.65, p = 0.07$, by independent-samples one-tailed t-test), but was supported for f0 ($t(19) = 3.38, p = .002$), indicating that speakers who had spent longer than 6 months in an English-speaking country used pitch less (i.e. in the predicted, more native-like direction) in producing their L2 English voicing contrast than the speakers who had not. Although the result for f0 is suggestive of the role of experience in relative cue use in production, our subject group is not appropriately large and demographically balanced to make strong claims about the effect of experience. See Kong and Yoon (2013) for a direct test of this question.

the types are all distinct when taking into account both VOT and f0. This pattern held for all individual speakers: none have any stops overlapping on both f0 and VOT duration, even for Korean aspirated and English voiceless stops, categories that have been suggested in previous work to show assimilation (e.g. Schmidt 1996; Kang and Guion 2006). However, it is important to keep in mind that the Korean and English target stops occurred in different prosodic positions and are therefore not directly comparable.




(a) f0 by VOT

(b) Closure duration

Figure 3: Comparison of production values for English and Korean stops on the dimensions of (a) f0 and VOT duration and (b) closure duration. Error bars for f0 and VOT represent one standard deviation; the box plots for closure duration show the interquartile range (outliers $+/-2SD$). Graphs are based on the mean values from each speaker; all values are standardized such that each speaker's production values across both languages for each dimension are centered around zero.

### 2.3 Interim Discussion: Production data

The Korean production values found here are in line with recent work (e.g. Lee and Jongman 2012) in terms of VOT and f0: Korean stop categories all differed from one another on both dimensions, although values for VOT duration were heavily overlapping for aspirated and lenis stops. VOT duration was the most reliable dimension distinguishing aspirated from fortis stops, while f0 most reliably distinguished aspirated from lenis stops; the lenis-fortis contrast was well-separated on both dimensions. Speakers also distinguished all three pairwise contrasts using closure duration (most reliably for the fortis-lenis contrast), showing the same pattern as earlier work (Silva 1992; Kim 1994, but see Cho and Keating 2001). For English, both VOT and f0 were very reliable indicators of category membership, and the heavy use of f0 in distinguishing speakers' productions of the two types of stops stands in contrast to the productions of native English speakers, in which f0 is only weakly predictive of stop category membership (Löfqvist et al. 1989; Shultz et al. 2012). Speakers' consistent use of two separate cues to define a single two-way contrast leads to redundancy in how the contrast is specified; in other words, if voiceless stops are produced with both long VOT and high f0, and voiced stops are produced with both short VOT and low f0 (as is generally the case, see Figure 1), knowing either the value of VOT or f0 would allow for an accurate categorization of the stop. However, this redundancy also leads to "undefined" areas of the acoustic space; for example, the question of how a stop with *long* VOT and *low* f0 would be classified is not predicted by the production data. In principle, listeners could make different choices in how to categorize the remainder of the acoustic space. We explore this possibility in the following sections.

Speakers also appeared to use closure duration to distinguish the contrast in production; a main effect was not present on a group level, but the subset of participants who did not produce a prosodic break before the target syllable did show significantly longer closure durations for voiceless than for voiced stops. Although the five stops (Korean and English) were produced in distinct areas of the acoustic space by all speakers, direct cross-language phonetic category comparison is not possible because the prosodic environments for target sounds were not identical in the two languages.

## 3 Experiment 2: Perception

Experiment 2 turns to the same native Korean speakers' perception of Korean and English word-initial stop contrasts, and in particular, the relative influence of the three acoustic dimensions of VOT, f0, and closure duration in categorization of the contrasts.

### 3.1 Methods

*Stimuli:* Perception stimuli were created by recording natural productions of stop-initial syllables in the context of carrier sentences (English "Click on *pa*," Korean [tʃigɨm *pa*-ɾɨl nuɾisejo], "Now click on *pa*")[5]. Sentences and target syllables were recorded by one female native speaker of each language. The target syllables were then modified using Praat to create a series of stops covarying in VOT, f0 and closure duration, resulting in a set of stimuli spanning a three-dimensional acoustic space for each language: seven steps of VOT by seven steps of f0 by three steps of closure duration (the ranges for each dimension are given in Table 5). A larger number of steps were used for VOT and f0 in order to get a more fine-grained picture of the cues that have been shown in previous work to be more primary cues in Korean, whereas closure duration is only a secondary cue in both languages. Stimuli were created as follows: first, the baseline token was a lenis stop for Korean and a voiceless stop for English (these were determined to be the best baseline tokens in pilot work). VOT duration and f0 were modified using the "To Manipulation…" function in Praat; closure duration was manipulated by adding increasing amounts of silence between the preceding word in the carrier phrase and the target syllable. The onset f0 was set to vary in the range given above, remained at a stable frequency for one-third of the vowel duration, then fell linearly to a final f0 of 140 Hz at the end of the vowel. Therefore, all of the stimuli had the same final f0; only the onset varied. The endpoints of the f0 and closure duration series were chosen by taking 0.25 standard deviations above and below the maximum and minimum values from the stop contrast in pilot production work from the same speaker for each language. The VOT range was chosen to be comparable to previous studies, spanning the native English phonetic category boundary. The target syllables were then embedded into the language-specific carrier sentences.

#### 3.1.1 Participants

The same group of participants from the production experiment participated in the perception experiment for a given language in the same session.

#### 3.1.2 Procedure

The Korean and English sessions were completed on two separate days, and oral and written instructions were provided in the target language on each day. On each day, the perception task

[5]Using a real word would have been more parallel to the production task. However, we chose to use a nonsense syllable, as opposed to a real minimal pair or triplet from each language, in order to avoid word-specific lexical effects.

| Language | Parameter | Range | Step size | Number of steps |
|---|---|---|---|---|
| **Korean** | VOT | 0 - 84$ms$ | 14$ms$ | 7 |
| | f0 | 215 - 334$Hz$ | 19$Hz$ | 7 |
| | closure | 25 - 125$ms$ | 50$ms$ | 3 |
| | | | | **Total = 147 stimuli** |
| **English** | VOT | $-20$ - 40$ms$ | 10$ms$ | 7 |
| | f0 | 141 - 225$Hz$ | 14$Hz$ | 7 |
| | closure | 40 - 120$ms$ | 40$ms$ | 3 |
| | | | | **Total = 147 stimuli** |

Table 5: Range, number of steps, and step size for each parameter manipulated in the acoustic series of stimuli.

(Experiment 2) for a given language was completed after the production task (Experiment 1) in the same language. The experiment was presented using PsychoPy (Peirce 2007). Listeners were presented with a forced-choice task in which they heard language-specific stimuli (embedded in carrier sentences from each language) encompassing the full range of the three-dimensional acoustic space described above. For each stimulus, the listeners were asked to choose which of two (English) or three (Korean) sounds best represented what they heard by pressing a specified key. The listeners heard four randomized blocks of the complete stimulus set (147 tokens x 4 blocks = 588 total tokens). The task took about 30 minutes.

#### 3.1.3 Statistical analysis - Group results

We used logistic regression (as implemented with the *multinom* function in the *nnet* package in R) to analyze Koreans' perceptual patterns in their perception of Korean and English stops as a function of VOT duration, f0, and closure duration. A multinomial (as opposed to binary) function was used for the Korean three-way stop contrast in order to best model the task (the model takes into account the fact that there are three response choices). Aspirated was set as the default level, and the statistical results show the extent to which a change in each acoustic dimension elicits more Fortis or Lenis responses, respectively. A binary model was used for the English two-way contrast. The goodness-of-fit was similar between models including and excluding interactions, we therefore report the models without interactions in order to aid interpretability and maintain comparability with the production data, which do not include interactions between factors. All variables (VOT, f0, and closure duration) were standardized prior to analysis such that the resulting beta-coefficients for a given variable show the change in log odds of the relevant response given a single standard-deviation increase in that variable.

#### 3.1.4 Statistical analysis - Individual results

Individual participants' cue weights were computed via separate logistic regression analyses for each speaker with VOT duration, f0, and closure duration as predictors of perceived stop category (see Nearey (1997); Morrison (2005); Morrison and Kondaurova (2009) for discussion of use of logistic regression coefficients as metrics for perceptual cue weights). In order to keep the individual perceptual weights as comparable as possible to the those calculated via LDA for production, we performed three separate two-way analyses to model the three-way contrast (instead of the multinomial regression used for the group data), and interaction terms were again excluded. The beta-coefficients from each model were taken as an approximation of a given subject's reliance on

a given cue for the relevant comparison, and since they are based on standardized values for each dimension, can be compared to one another in order to determine the relative weighting of each dimension in predicting response patterns.

### 3.2 Results from Experiment 2: Perception

#### 3.2.1 Perception of L1 Korean stops

Heat plots showing group response patterns across each pair of cues for each stop type are shown in Figure 4. Results of the regression analyses are shown in Table 6, and a graph of predicted values across different levels of VOT and f0, overlaid with participants' response data, is shown in Figure 5.

The graphs in Figure 4 show listeners' categorization patterns, plotted across each combination of VOT, f0, and closure duration for each stop type separately. Each cell represents one stimulus, with darker cells representing a larger proportion of responses for a given category. Based on visual inspection of these graphs, listeners' Aspirated responses cluster in the acoustic space defined by long VOT and high f0, while the distributions do not appear to differ based on closure duration. Low f0 appears to serve as the major cue to Lenis classification, with shorter closure durations eliciting more Lenis responses. Stimuli with mid-to-high f0 and above all, short VOT, were considered Fortis, with a tendency for longer closure durations to elicit more Fortis responses.

The statistical results in Table 6 show that all three variables (VOT duration, f0, and closure duration) significantly influence listeners' response patterns. F0 is the most predictive dimension in separating Lenis responses from Aspirated, with lower f0 eliciting more Lenis responses; however, VOT and closure duration also secondarily influence perception of this contrast, with shorter VOTs, and, to a lesser extent, shorter closure durations, signalling the Lenis category. On the other hand, VOT is the strongest predictor of Fortis, as opposed to Aspirated, stops, with shorter VOTs signalling the Fortis category, while f0 and closure duration play smaller roles, with lower f0 and longer closure duration eliciting more Fortis responses.

| | Fortis (vs. Aspirated) | | | Lenis (vs. Aspirated) | | |
|---|---|---|---|---|---|---|
| Effect | $\beta$ | $z$ | $p <$ | $\beta$ | $z$ | $p <$ |
| Intercept | −1.15 | −16.88 | .001 | −0.50 | −11.03 | .001 |
| VOT | −4.24 | −46.41 | .001 | −1.56 | −23.64 | .001 |
| f0 | −0.21 | −4.12 | .001 | −4.26 | −43.83 | .001 |
| closure | 0.01 | −2.58 | .05 | −0.46 | −10.09 | .001 |

Table 6: Native Koreans' perception of L1 Korean stops: Beta-coefficients, z-scores, and p-values from the multinomial logistic regression model of perception of the Korean stop contrast. The $\beta$-coefficient for a given variable shows the increase in log odds of the Fortis or Lenis (as compared with the default Aspirated) response for each one-deviation increase in value for that variable.

Finally, Figure 5 allows for comparison of listeners' responses with the response curves predicted by the statistical model for low, mid, and high values of f0 across the VOT range (for purposes of visual clarity, f0 was collapsed into three levels, with responses collapsed over all values for closure duration). The relatively close fit of the curves to the actual responses shows that the model provides a good estimate of the actual responses and reinforces the statistical patterns discussed above; in particular, the graph shows the pattern that stimuli with low f0 are mainly predicted to be classified as Lenis, while the classification of stimuli with mid and high f0 depends primarily on VOT (those with high VOT being classified as aspirated, and low VOT being classified as Fortis).

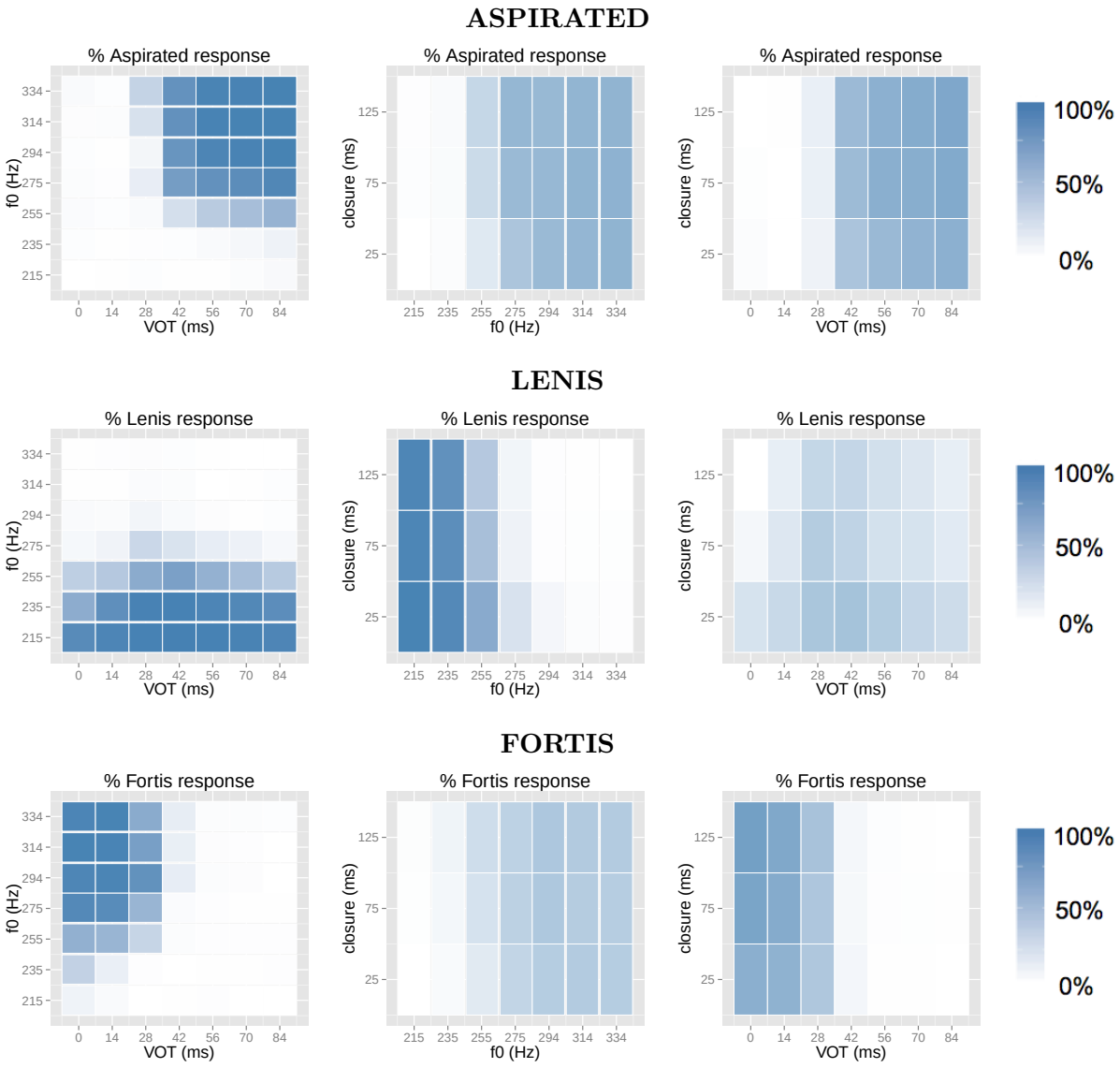



Figure 4: Heat plots of listeners' categorization of Korean word-initial stops, shown for each stop type plotted across each combination of the three acoustic dimensions manipulated in this experiment. Each cell represents one stimulus in the series, and the darkness of the cell represents the percentage response for each category given in a forced-choice task; for example, for the "Fortis" graphs, the darkest cells elicited the highest portion of Fortis response, while the white cells elicited the lowest.

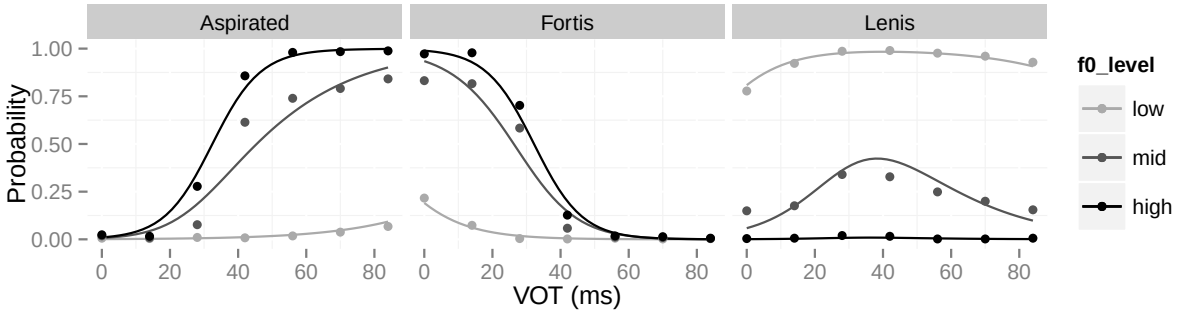
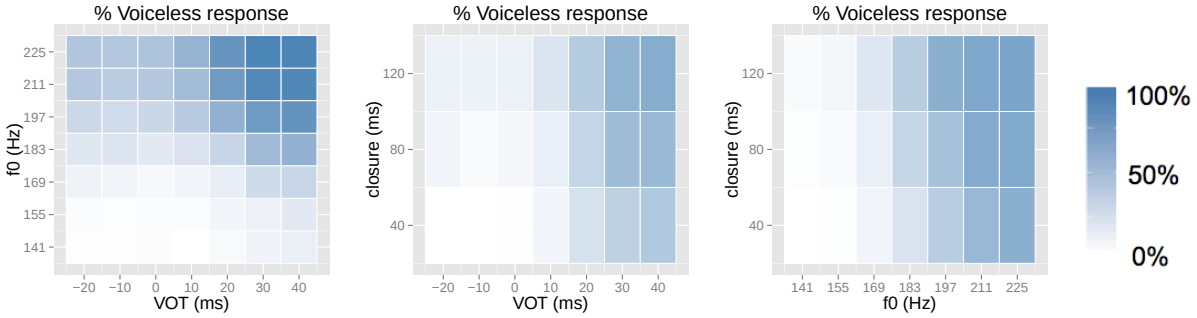



Figure 5: Predicted logit curves and average response data (points) for perception of the Korean three-way contrast in terms of VOT and f0 based on the multinomial logistic regression model (Table 6). The y-axis of each panel represents probability of a given response (Aspirated, Fortis, or Lenis). Points showing listeners' actual response data are overlaid.

### 3.2.2 Perception of L2 English stops

Heat plots showing L1 Korean listeners' responses to the (L2) English stimuli are shown in Figure 6, and statistical results of the group logistic regression analysis on categorization of English stops are given in Table 7.




(a) f0 by VOT (b) Closure by VOT (c) Closure by f0

Figure 6: Heat plots of Koreans' categorization of the English stop contrast, shown plotted across each combination of the three acoustic dimensions. Each cell represents one stimulus, and the darkness of the cell represents the percentage "voiceless" response in a forced-choice task; the darkest cells elicited 100% 'pa' response, while white cells elicited 100% 'ba.'

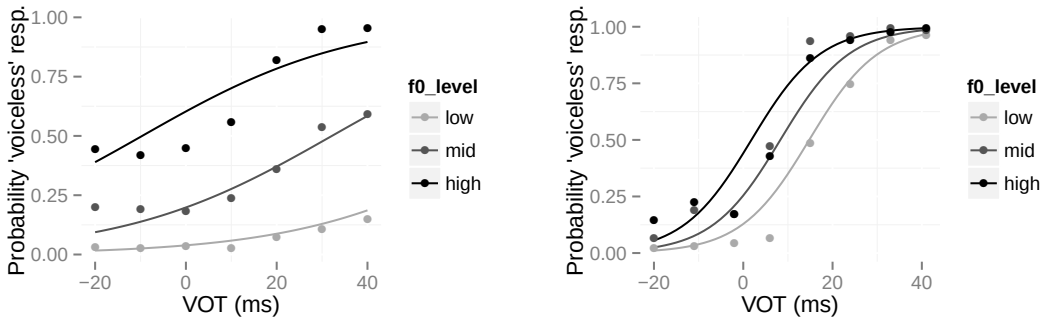
Visual inspection of the plots in Figure 6, which indicate the proportion "voiceless" (vs. "voiced") response in terms of the darkness of the cells, show that overall, only stimuli with *long* VOT and *high* f0 were consistently classified as voiceless, while only stimuli with *short* VOT and *low* f0 were consistently classified as voiced. The other two quadrants of the space show intermediate response values. There is also a tendency for stops with longer closure duration to be classified as voiceless. The results of the regression analysis in Table 7 confirm these patterns, showing that all three acoustic dimensions had a significant effect on choice of stop type, with longer VOT, higher f0, and longer closure duration all eliciting more Voiceless responses.

The response curves predicted by the model are shown in Figure 7a, overlaid on listeners' averaged response data (as with the Korean data above, collapsed into three levels of f0 and collapsed

| | Voiced vs. Voiceless | | |
|---|---|---|---|
| Effect | $\beta$ | z | $p <$ |
| Intercept | −1.02 | −33.87 | .001 |
| VOT | 0.92 | 29.38 | .001 |
| f0 | 1.52 | 49.42 | .001 |
| closure | 0.31 | 10.96 | .001 |

Table 7: Koreans' perception of L2 English stops: results of a binary logistic regression model. Coefficients represent the effect of a given factor on the probability of eliciting a "voiceless" (vs. "voiced") response.

over all levels of closure duration). For comparison, we provide a graph of data from native English listeners' responses to a similar set of stimuli in Figure 7b. These data come from Schertz et al. (submitted). The stimuli were created in the same way, but only VOT and f0 (not closure duration) were manipulated. The results are based on fewer responses per subject because the task was only a subset of the study. The control English listeners show a canonical response curve with high and low values of VOT eliciting voiceless and voiced responses, respectively, regardless of the level of f0, while f0 plays only a secondary role, shifting the curves slightly such that higher f0 elicits more voiceless responses. On the other hand, the response curves of the Korean listeners demonstrate the greater importance of f0 for these listeners: the endpoints of the VOT series do not show categorically different responses, but rather still depend on heavily on f0 (with high f0 eliciting voiceless and low f0 eliciting voiced responses).




(a) Native Korean/L2 English

(b) Native English comparison data

Figure 7: Predicted logit curves for perception of the L2 English stop voicing contrast by native Korean listeners (a) and native English listeners (b). Predictions are drawn from a binary logistic regression model with VOT, f0, and closure duration as predictors. The y-axis of each panel represents the predicted probability of a "voiceless" response, and points showing listeners' actual response data are overlaid.

### 3.2.3 Perception: Individual results

Beta-coefficients from individual regression models for the English contrast and the L1 Korean contrasts are given in Table 13 in the Appendix, and plotted in Figure 8. For L2 English data, individual heat plots are also given alongside the individual production graphs in the Appendix (Figure 11). For the Korean comparisons, the distribution of individual listeners' coefficients, as shown in Figure 8 looks fairly similar to the coefficients calculated from individual speakers'

productions (Figure 2), with greater use of VOT than f0 in distinguishing the fortis vs. aspirated contrast and greater use of f0 than VOT in distinguishing the aspirated vs. fortis contrast, as well as the fortis vs. lenis contrast (but with both VOT and f0 playing an important role). However, the relative importance of cues for the English contrast is the reverse of that found in production: f0 is a better predictor than VOT of listeners' categorization of stops, while VOT was a more reliable separator of their productions. Furthermore, there was again considerable variability between listeners, with f0 serving as the strongest predictor for twelve listeners, and VOT serving as the strongest predictor for the other nine. Below we consider to what extent this variability corresponds to the variability found in speakers' productions of the stops[6].

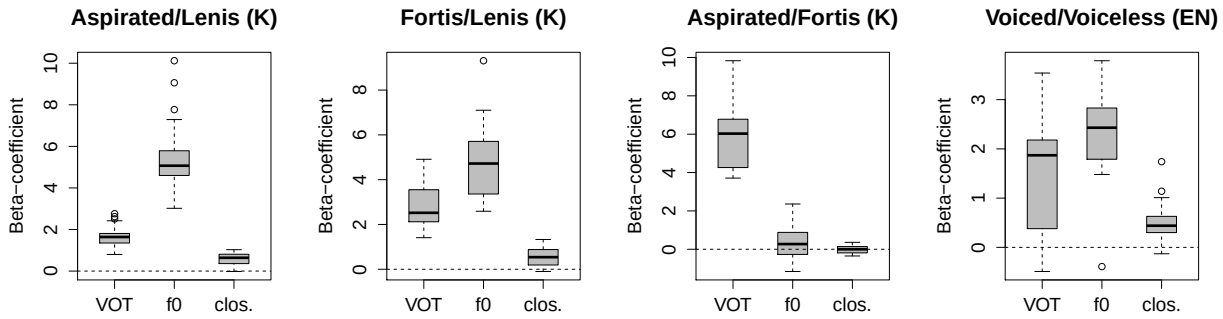
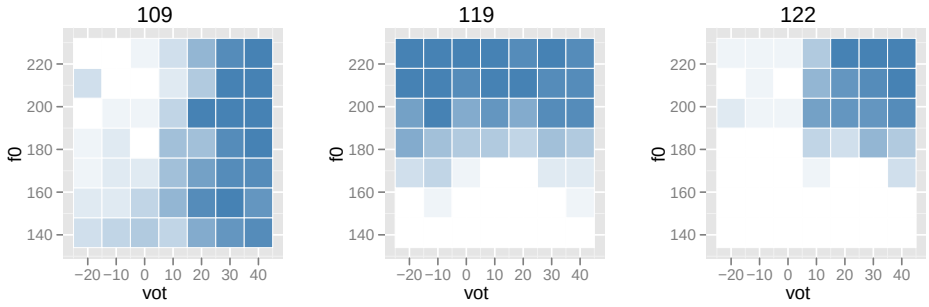



Figure 8: Predictability of listener responses based on each cue manipulated in Experiment 2, as estimated by the distribution of coefficients from individual logistic regression analyses for each speaker/contrast. Boxplots show the interquartile range (and any outliers beyond 2 standard deviations from the mean) of speakers' coefficients for each dimension. A larger coefficient indicates a larger influence of a given dimension on listeners' responses.

Heat plots of the individual L2 English perception results (Figure 11) show considerable qualitative variability in individuals' relative use of f0 and VOT, compared to the relative homogeneity of productions. To give an example of the range of variation, the individuals with the largest differences between perceptual weights for VOT and f0, as well as a subject who weighted both dimensions relatively equally, are shown in Figure 9. Therefore, the somewhat gradient-looking use of f0 and VOT seen in the plot of the group results (Figure 6) masks what appear to be, based on visual inspection of the plots, relatively sharp boundaries indicating more categorical cue-weighting strategies on an individual level. However, this variability is not random; in particular, the regions of the acoustic space used in production of the contrast were categorized in the same way by all listeners: stimuli with long VOT and high f0 were categorized as voiceless, while stimuli with short VOT and low f0 were categorized as voiced. Instead, the variability occurs in the areas of the acoustic space not defined by production patterns (i.e. long VOT paired with low f0 and short VOT paired with high f0); in other words, it appears that listeners use different strategies when classifying the regions of the acoustic space not used in production.

[6]Parallel to the similar analysis with production weights, we explored the hypothesis that listeners with more experience with English might have higher perceptual weights for VOT and lower weights for f0 than their counterparts with less English experience. However, independent samples one-tailed t-tests did not show significantly greater use of VOT for more experienced listeners ($t(20) = 0.4, p > .1$) or significantly less use of f0 ($t(20) = -.60, p > .1$).




(a) VOT (b) f0 (c) VOT and f0

Figure 9: Heat plots of the individuals with the largest differences between perceptual weights for VOT and f0 (a and b), as well as an individual with relatively equal weightings on the two dimensions (c). Cell darkness represents percentage "voiceless" response.

### 3.3 Experiment 2a: Stability of L2 perceptual weights across sessions

Given the considerable differences in perceptual cue weighting strategies found in Experiment 2, it is important to confirm that these individual patterns are stable across time: it is possible that the individual variability found in Experiment 2 does not actually capture stable independent strategies for distinguishing the stop contrast, but rather random noise in listeners' perceptual patterns. Twelve of the participants from the previous two experiments also participated in a follow-up experiment on a different day. Although the data were drawn from a different study with different goals, the first block of the experiment was identical in procedure to the the forced-choice perception task of Experiment 2, giving us the opportunity to compare the data from these two sessions in order to examine the stability of perceptual cue weighting strategies.

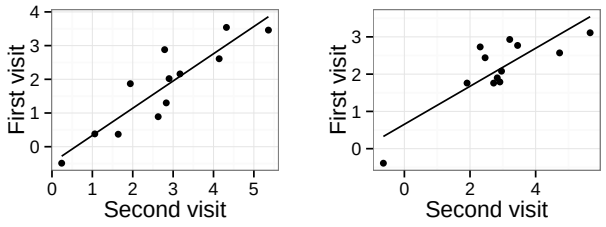
#### 3.3.1 Methods

*Stimuli:* The stimuli were created in the same way as the English stimuli in Experiment 2: listeners heard target syllables embedded in the English carrier sentence "Click on *pa*." A natural English voiceless stop was manipulated on the dimensions of VOT duration and f0, as in Experiment 2; closure duration was not manipulated for the stimuli for this experiment. The ranges for VOT duration and f0 were slightly different (f0: 160 to 240 Hz, as opposed to 141 to 225 Hz in Experiment 2; VOT duration: -20 to 50ms, as opposed to -20 to 40 in Experiment 2), and there were nine steps of each (as opposed to seven in Experiment 2), for a total of 81 distinct stimuli (the ranges and number of steps differed because of the nature of the rest of the experiment).

*Participants:* Twelve of the participants from Experiments 1 and 2 participated in a follow-up experiment, which took place between two weeks and a month later.

*Procedure:* The first block of the experiment (which is the only portion of the experiment discussed in the current work) was identical in procedure to Experiment 2: listeners participated in a forced-choice task ('pa' vs. 'ba'). In this block, listeners heard two repetitions of the 81 stimuli, for 162 trials per subject (as opposed to 588 in Experiment 2).

*Statistical Analyses:* We used univariate logistic regression models to obtain each listeners' perceptual weight, as approximated by the beta-coefficient of the regression model, for VOT duration and f0. We then performed one-tailed Pearson product-moment correlations to determine whether the individual weights were correlated across the sessions on each dimension.




(a) VOT weight (b) f0 weight

Figure 10: Correlation of perceptual weights for f0 and VOT duration across first and second visits (Experiments 2 and 2a). Each point represents a different participant.

#### 3.3.2 Results

Listeners' weights, as estimated by the beta-coefficients of the regression models across the two sessions for each acoustic dimension (VOT duration and f0) are plotted in Figure 10. There was a statistically significant and strong correlation for both dimensions (VOT: $r = .90, p < .001$, f0: $r = .84, p < .001$), indicating that listeners' reliance on each of the dimensions was relatively stable across experimental sessions.

## 4 Further analysis: Comparing cue use across production and perception

### 4.1 Predictability of group results across modalities

In order to compare the use of the three acoustic dimensions consistently across the two modalities, we used classification accuracy from Linear Discriminant Analysis (LDA) models of group data in each language and modality. As with the analyses used above for individual production data, these LDA models incorporated the three acoustic dimensions as predictor variables for stop types (in production) and listener responses (in perception). We then used the optimized model in each modality to classify the data. The classification accuracy of each model can therefore be thought of as a metric of how well the dataset can be separated using an optimized linear combination of the three dimensions.

Confusion matrices and classification accuracies for the three pairwise comparisons for Korean and English production and perception are shown in Table 8. For the production data, all models performed well (over 96% accuracy). For the models classifying the perception data, accuracy was lower for all pairwise comparisons than it was for production; however, there was much lower accuracy for English (79.1%) than for any of the Korean contrasts (all above 90%). This markedly lower accuracy quantifies the gradience or "fuzziness" seen in the graph of the English group perception data (Figure 6), as compared with the Korean perception data, or the production data from either language. If this lower accuracy stems from heterogenous cue weighting strategies, as is our interpretation, the lower accuracy measure should be driven by the stimuli where groups should differ (i.e. Quadrants II and IV). In order to test this, we built an additional model with only the canonical tokens that were expected to be classified in the same way by all listeners, regardless of relative cue-weighting strategy (i.e. the long-VOT, high-f0 stimuli in Quadrant I and the short-VOT, low-f0 stimuli in Quadrant III). Classification accuracy of the model trained on only these canonical stimuli was 90.4%, more comparable to the models of L1-Korean perception.

PRODUCTION

| | For. vs. Len | | | Asp. vs. For. | | | Asp. vs. Len. | | | Voiced vs. Voiceless | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classification: | | For. | Len. | | Asp. | For. | | Asp. | Len. | | Vc. | Vless. |
| Type: | For. | 734 | 2 | Asp. | 758 | 38 | Asp. | 787 | 9 | Vc. | 1038 | 11 |
| | Len. | 7 | 734 | For. | 9 | 727 | Len. | 2 | 739 | Vless. | 36 | 1019 |
| **Accuracy:** | | | **99.3%** | | | **96.9%** | | | **99.3%** | | | **98.1%** |

PERCEPTION

| | For. vs. Len | | | Asp. vs. For. | | | Asp. vs. Len. | | | Voiced vs. Voiceless | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classification: | | For. | Len. | | Asp. | For. | | Asp. | Len. | | Vc. | Vless. |
| Choice: | For. | 2892 | 433 | Asp. | 4243 | 248 | Asp. | 4054 | 437 | Vc. | 6900 | 1167 |
| | Len. | 233 | 4299 | For. | 153 | 3172 | Len. | 276 | 4256 | Vless. | 1416 | 2865 |
| **Accuracy:** | | | **91.5%** | | | **94.5%** | | | **92.1%** | | | **79.1%** |

Table 8: Confusion matrices showing classification errors of discriminant analysis models for each pairwise comparison of the Korean stop contrast and the English stop contrast in production and perception. Models used all three dimensions (aspiration, f0, and closure duration) to predict stop type in production, and choice of stop type in perception.

### 4.2 Individual production-perception correlations

It might be expected that the informativeness of a given dimension in a given speaker's productions would correlate with the weight given to that dimension in perception by the same speaker. We examined participants' cue weighting strategies across modalities via correlation analyses, using the individual LDA coefficients and beta-coefficients as approximations of each individual's use of cues in production and perception respectively; that is, whether the individual perceptual weights could be predicted by the production weights. However, there was not a significant positive correlation between the weights in the two modalities (using Pearson product-moment correlation) for any of the dimensions (Table 9).

| | Voiced vs. Voiceless | | Aspirated vs. Lenis | | Lenis vs. Fortis | | Aspirated vs. Fortis | |
|---|---|---|---|---|---|---|---|---|
| | $r$ | $p$ | $r$ | $p$ | $r$ | $p$ | $r$ | $p$ |
| VOT | $-.01$ | $> .1$ | $.04$ | $> .1$ | $.03$ | $> .1$ | $.04$ | $> .1$ |
| f0 | $.02$ | $> .1$ | $-.15$ | $> .1$ | $.40$ | $= .07$ | $-.15$ | $> .1$ |
| closure | $.34$ | $> .1$ | $-.26$ | $> .1$ | $.02$ | $> .1$ | $-.25$ | $> .1$ |

Table 9: Results of correlation analyses of perception vs. production weights for each contrast.

## 5 Discussion and Conclusion

### 5.1 Summary of results

The Korean participants' production and perception of their native stop contrast largely mirrored recent work on cue weighting in the Seoul Korean stop contrast (e.g. Kang and Guion 2006; Lee and Jongman 2012; Lee et al. 2013). VOT duration was the most important cue for the aspirated vs. fortis contrast, both in production and perception, and closure duration was a secondary cue in both modalities. Speakers' f0 values differed significantly between the two stop types; however, f0 did not influence listeners' classifications in perception. For the aspirated vs. lenis contrast, f0 was the most important dimension in both perception and production, with VOT and closure

duration serving as secondary cues in both modalities, in line with recent work showing that the aspirated vs. lenis contrast, differentiated in the past by VOT duration, is now primarily cued by f0 for younger speakers of the Seoul dialect (Silva 2006; Lee and Jongman 2012; Lee et al. 2013; Kang 2014). All three dimensions contributed to both the production and perception of the fortis vs. lenis contrast in both modalities; in production, VOT and f0 were the most reliable indicators, followed by closure, although all were fairly reliable, while perceptual results showed the strongest effect of f0, followed by VOT, followed by closure duration.

The same Korean participants used both VOT, and to a somewhat lesser extent, f0, very reliably to distinguish their L2 English stop contrast in production (cf. Kang and Guion 2006). In perception, f0 was the strongest cue to the distinction, followed by VOT (cf. Kim 1994). In both perception and production, the use of f0 is much stronger than in native English listeners, where f0 is only a very weak cue to the distinction (cf. Francis et al. 2008; Kingston et al. 2008; Llanos et al. 2013). Closure duration influenced non-native (Korean) listeners' classification of English stops and also distinguished speakers' productions of the contrast when it was not conflated with a perceptible prosodic boundary before the target word. Although individual strategies differed in both perception and production, different speakers' relative use of the cues in production was more homogenous than in perception: while stop type in production was highly predictable when the three dimensions were taken into consideration (by a model based on a linear discriminant analysis), listeners' choice of stop type was much less predictable. This unpredictability, or "fuzziness" in perception, also apparent in the graphs of group perception data, appears to be due to different individual strategies in perceiving the stop contrast: some users relied primarily on VOT duration, some on f0, and some used both to an equal extent, and these individual strategies were stable across sessions. Production patterns were not predictive of cue-weighting strategies in the perception task: individual perception and production weights for both the English and the Korean contrasts were not correlated.

### 5.2 Individual variability in L2 perception

The individual variability present in the L2 perception data is striking, especially given the fact that more listeners defined the contrast using pitch than VOT in perception of the English contrast, even while showing a highly reliable VOT distinction in production. This asymmetry may be in part due to the fact that the Korean and English stop contrasts inhabit different areas of the acoustic space, even though they are both cued by the same acoustic dimension (VOT). For example, Cho and Keating (2001, 2009) showed that in a comparable utterance-initial position, the average VOT was about 40 ms for an English /t/, but about 60 ms for the Korean counterpart (produced by native speakers of each language); and Kim and Cho (2013) demonstrated that Korean listeners required a significantly longer VOT for categorizing English aspirated stops than native English listeners. Recall that Korean speakers produced both Korean aspirated stops and English voiceless stops with substantially long VOT duration, which averaged 86 versus 95 ms, respectively. In other words, Korean speakers appear to have produced English voiceless stops by recruiting the same portion of acoustic space used for their native stop contrast. The finding that native Korean listeners did not reliably use VOT duration to distinguish the English stop contrast in the current perception study may therefore be attributable to the fact that the English stimuli did not have "prototypical" (for Korean listeners) values on this dimension. This implies that L1 influence on L2 cue weighting is further constrained by the extent to which the specific region of the acoustic space utilized in the native language is shared by the L2.

If we assume, based on the native Korean participants' production patterns, that good exemplars of English voiceless stops have high f0 and long VOT, while good exemplars of voiced stops have

low f0 and short VOT, then the contrast is in a sense "overspecified," with two dimensions defining a single binary contrast. This redundancy in principle provides multiple, equally "correct" options for a language learner trying to determine the importance of the two cues: one listener could choose to rely primarily on VOT, another primarily on f0, and a third could pay attention to both, and all three listeners would correctly categorize the prototypical stops. These differences in cue-weighting would only appear in categorization of non-prototypical sounds; that is, sounds with "mismatched" VOT and f0 (e.g. high f0 and *short* VOT, or low f0 and *long* VOT). The results of the current work suggest that listeners do indeed make different choices (albeit not necessarily consciously). While all listeners converged on "voiceless" responses for perceptual stimuli in Quadrant I (long VOT + high pitch) and "voiced" responses for stimuli in Quadrant III (short VOT and low pitch), individuals differed in how they categorized stimuli in Quadrants II and IV, precisely those areas of the acoustic grid that are good exemplars of neither voiced nor voiceless stops. We might have expected random responses in these quadrants, resulting in "fuzzy" boundaries, which is a common description of the results of L2 cue weighting studies (e.g. Morrison 2007; Escudero et al. 2009). Consistent with this expectation, the group perception data does not show clear category boundaries (Figure 6). However, by contrast, the individual graphs show sharper boundaries (Figure 11), suggesting that the group fuzziness in the current data results from the combination of different individual patterns, as opposed to consistent uncertain classification. The different groupings or perceptual strategies (groups using primarily pitch, primarily VOT, or both dimensions) found in the current data thus appear to represent three different options used by listeners to classify sound categories in the absence of complete information across the acoustic grid. The fact that these differences are stable across experiment sessions, as opposed to being random noise in the data, indicates that these cue weightings are a structured part of the L2 sound system.

Regardless of the source(s) of these individual differences, recent work has shown that variability in phonetic cue-weighting strategies can also drive differences in L2 learning and adaptation patterns: Chandrasekaran et al. (2010) found that the extent of improvement in a Mandarin tone training paradigm was partially attributable to listeners' use of pitch height vs. direction in their initial perception of the contrast, while Wanrooij et al. (2013) (L1 Spanish/L2 Dutch) and Schertz et al. (submitted) (L1 Korean/L2 English) showed that the same L2 perceptual learning task elicited different adaptation patterns based on initial cue-weighting strategies. Therefore, differences such as those examined in the current work have relevance beyond phonetic categorization tasks and may be important to consider in the study of phonetic learning and adaptation mechanisms.

### 5.3 Native language influence on L2 production and perception

The current production and perception results converge to demonstrate that both VOT and f0 play an important role overall in native Koreans' representation of the L2 English stop distinction. This heavy reliance on f0 (relative to that of L1 English speakers) most likely results from the use of f0 as a primary cue in the native Korean stop contrast. Production patterns fit in with the prediction of SLM that speakers will dissimilate L2 phonetic contrasts within the relevant acoustic space (Flege 2007); that is, since the "relevant acoustic space" for Korean listeners perceiving stops appears to be the space encompassing the dimensions of f0, VOT, and closure duration, they would be expected to try to differentiate their L2 stop contrast along all three of these dimensions. Another account of the production patterns may be that (all) listeners are sensitive to distributional regularities in the input, but these general sensitivities are modulated by native language biases, such that more importance is given to the dimensions that are used as primary cues in the L1 (see Holt and Lotto (2006) for discussion of the interplay between these factors).

Previous work has suggested that English voiceless stops are assimilated to the Korean aspirated stop category by late bilingual Korean speakers (Schmidt 1996; Kang and Guion 2006; Park and de Jong 2008). Direct comparison of the languages is not possible in the current work because the series used for the perception experiments encompassed different acoustic spaces for Korean and English. Furthermore, the three-to-two mapping of Korean to English stops makes it such that even if the two English categories were assimilated to two different Korean categories, it would not be apparent from the results of a forced-choice experiment such as the one used in the current work (because the listeners would be forced to choose one of the English categories for the "unused" Korean category). The results of some individual listeners suggest that the perceptual cues used to define English voiceless stops are the same as those used to define Korean aspirated stops (i.e. those participants who required both high f0 and long VOT for "voiceless" responses). However, this is not the case for all listeners. Listeners who showed a more unidimensional division of the acoustic space (i.e. those who made the distinction primarily based on f0 or primarily based on VOT) cannot be basing these judgments solely on an "English voiceless = Korean aspirated" mapping. Korean listeners' perception of English stops therefore does not fit into a straightforward account of L1-L2 category assimilation.

Most explanations for variability in L2 listeners' use of phonetic cues depend on either the acoustic relationship or distance between the L1-L2 categories (e.g. Best 1995; Flege 1995), the L2 listeners' proficiency (e.g. Flege et al. 1996; Bohn and Flege 1997; Baker and Trofimovich 2005), or the specific dialect of the L2 being acquired (Escudero and Boersma 2004). In the case examined in the current work, none of these factors appears to account for the individual variability found in the perception data. Instead, this seems to be a case in which listeners have multiple "correct" ways to define the English voicing contrast in perception, since it is reliably separated by two different acoustic cues in production. Theories of L2 category acquisition must therefore be able to account for how listeners settle on a mapping when there are multiple, equally viable cue weighting strategies for a given sound contrast.

### 5.4 Production-perception interface

As discussed above, there appears to be much more variability in native Korean listeners' perceptual cue weights for the L2 English stop contrast than there is in their productions of the English contrast. In production, almost all speakers produce a reliable difference in VOT across the two stop types. On the other hand, the perception data are marked by divergent patterns, with some listeners dividing the acoustic space primarily on the dimensions of f0 or VOT, and others showing gradient effects of both or requiring both long VOT and high pitch to classify a stop as voiceless. The poor performance of a classifier based on the perceptual data as compared to the production data provides quantitative confirmation that the production values were more homogenous than the perception patterns.

The differences found in perception do not appear to be directly linked to differences in production in either the L1 or the L2 comparisons, at least in terms of the metrics used for comparison in the current work. There are reasons to be cautious about generalizing this lack of finding to a broader relationship between perception and production. First, differences between the two tasks, particularly as they are implemented in the current work, may partially account for the lack of correlation. Most notably, the range of the acoustic space used for the English perception experiment was not isomorphic to the range used by the speakers for English production (although the ranges for the Korean tasks, which similarly failed to show a perception-production correlation, were comparable); furthermore, while real words were used for production, nonwords were used for perception stimuli. Another complicating factor was the homogeneity of production values,

with most speakers showing highly reliable use of both f0 and VOT, and some speakers completely separating the categories on both dimensions. This may have constrained the variability such that a correlation effect would be difficult to find.

More generally, this work also raises the question of whether the metrics standardly used to define cue "use" in perception and production are (conceptually) comparable. Most analyses of use of multiple cues in production rely on read speech tasks, allowing speakers to use any dimensions they choose in an unconstrained manner (e.g. Flege et al. 1997; Dmitrieva et al. 2015). On the other hand, perceptual cue weights are usually quantified via forced-choice tasks, such as the one used in the current work (e.g. Flege et al. 1997; Kondaurova and Francis 2008; Llanos et al. 2013). It may be the case that differences tend to arise primarily in the "ambiguous" region of the acoustic space (i.e. the unused regions of the production space), as they did in the current work, whereas there may be much less variability in the regions of the perceptual space that are actually defined in production. Therefore, the disparity between the sorts of tasks generally used to quantify cue use in the two modalities may contribute to the difficulty in finding a relationship between production and perception, even if this relationship does exist. A serious search for a perception-production link on an individual level may require more subtle or creative tasks and metrics for calculating cue weights.

Nevertheless, the patterns in the data demonstrate independence between perception and production patterns in both the L1 and the L2, at least in terms of the acoustic space examined in this work. Different degrees of informativeness on a given dimension in a speaker's productions do not predict how much weight listeners will give to that dimension in perception: a speaker who shows considerable overlap in f0 values for voiced and voiceless stops can make the perceptual distinction between the two categories entirely based on pitch (e.g. S105). Conversely, differences in perception patterns do not imply differences in production: the three participants whose perceptual results look categorically different in Figure 9 all show relatively similar production patterns: these speakers' categories are almost perfectly separated on f0 and VOT independently. The same is true for the secondary cue of closure duration. Some participants' individual values were relatively informative of category membership; however, as shown by the null effect in the correlation analysis, these production patterns were not consistent across the same participants' perception. For example, S116, whose closure durations distinguished the stop contrast highly reliably in production, showed almost no effect of closure duration in perception. In sum, the data therefore are not consistent with a strict view that individuals' use of various acoustic dimensions in production should predict the extent to which they use these cues when categorizing ambiguous stimuli, in line with similar recent work on both L1 (Shultz et al. 2012; Idemaru et al. 2012, but cf. Newman 2003; Perkell et al. 2004) and L2 (Bohn and Flege 1997). The lack of a perception-production correlation on these tasks also calls into question whether perceptual cue weights can be predicted from distributional information, as is assumed in models of speech perception positing a strong perception-production link (e.g. Liberman and Mattingly 1985; Fowler 1986), and if not, how and why these perceptual weights emerge.

### 5.5 Conclusion

The current work examined how speakers and listeners produce and perceive a non-native contrast that relies on fewer acoustic cues than their native contrast, namely native Koreans' production and perception of their L2 English stop contrast, as compared with their native Korean 3-way stop contrast. L2 production patterns were quite homogenous, with participants using both f0 and VOT duration to distinguish the English contrast in production. This reliable use of two cues for a binary contrast provided multiple options for which cue(s) to rely on in perception, as shown

by variable response patterns on a forced-choice task with stimuli varying across all dimensions. These individual differences in relative use of f0 and VOT as cues to the contrast were stable across experimental sessions, demonstrating that they constitute an internalized part of the L2 grammar. There was no correlation between individuals' use of any of the three cues across production and perception in either the L1 and the L2, suggesting some degree of independence between the modalities. The results from this work demonstrate how phonetic cue use in the L1 can both influence use of acoustic cues in a systematic way (as shown by speakers' consistent use of f0 in productions of the L2 contrast) and introduce individual variability (by providing listeners with multiple "correct" options for perceptual cue weighting).

**Acknowledgements**
The authors would like to thank Daejin Kim for his help running participants, Jae-Hyun Sung for recording stimuli, as well as Miquel Simonet, three anonymous reviewers, and the Associate Editor for helpful suggestions and feedback. This work was supported by NSF EAPSI grant #1311026 and NIH-NIDCD grant #R01DC004674.

## Appendix

| **Lenis** | | | **Fortis** | | | **Aspirated** | | |
|---|---|---|---|---|---|---|---|---|
| [pal] | 발 | 'foot' | [p*aŋ] | 빵 | 'bread' | [p$^{h}$al] | 팔 | 'arm' |
| [pi] | 비 | 'rain' | [p*im] | 삠 | 'sprain' | [p$^{h}$in] | 핀 | 'pin' |
| [pul] | 불 | 'fire' | [p*ul] | 뿔 | 'horn' | [p$^{h}$ul] | 풀 | 'grass' |
| [tal] | 달 | 'moon' | [t*al] | 딸 | 'daughter' | [t$^{h}$al] | 탈 | 'mask' |
| [taŋ] | 당 | 'party' | [t*aŋ] | 땅 | 'earth' | [t$^{h}$aŋ] | 탕 | 'hot bath' |
| [tem] | 댐 | 'dam' | [t*e] | 때 | 'time' | [t$^{h}$e] | 테 | 'rim/hoop' |
| [tʌk] | 덕 | 'virtue' | [t*ʌk] | 떡 | 'rice cake' | [t$^{h}$ʌk] | 턱 | 'chin' |
| [tɯk] | 득 | 'profit' | [t*ɯʃ] | 뜻 | 'meaning' | [t$^{h}$ɯl] | 틀 | 'housing/frame' |
| [ke] | 개 | 'dog' | [k*e] | 깨 | 'sesame' | [k$^{h}$e] | 캐 | 'digs out' |
| [kom] | 곰 | 'bear' | [k*on] | 꼰 | 'braided' | [k$^{h}$oŋ] | 콩 | 'beans' |
| [kol] | 골 | 'goal' | [k*ol] | 꼴 | 'form/figure' | [k$^{h}$o] | 코 | 'nose' |
| [kʌm] | 검 | 'sword' | [k*ʌm] | 껌 | 'chewing gum' | [k$^{h}$ʌp] | 컵 | 'cup' |
| [kuk] | 국 | 'soup' | [k*um] | 꿈 | 'dream' | [k$^{h}$uŋ] | 쿵 | 'crash' |

Table 10: Korean production stimuli (transcription, Korean orthography, and English glosses)

| **Voiceless** | **Voiced** | **Voiceless** | **Voiced** | **Voiceless** | **Voiced** |
|---|---|---|---|---|---|
| pad | bad | tip | dip | came | game |
| pack | back | test | desk | coat | goat |
| pest | best | top | dot | cold | gold |
| pig | big | teen | dean | cave | gave |
| pill | bill | ten | dent | coal | goal |
| path | bath | tune | dune | coast | ghost |

Table 11: English production stimuli

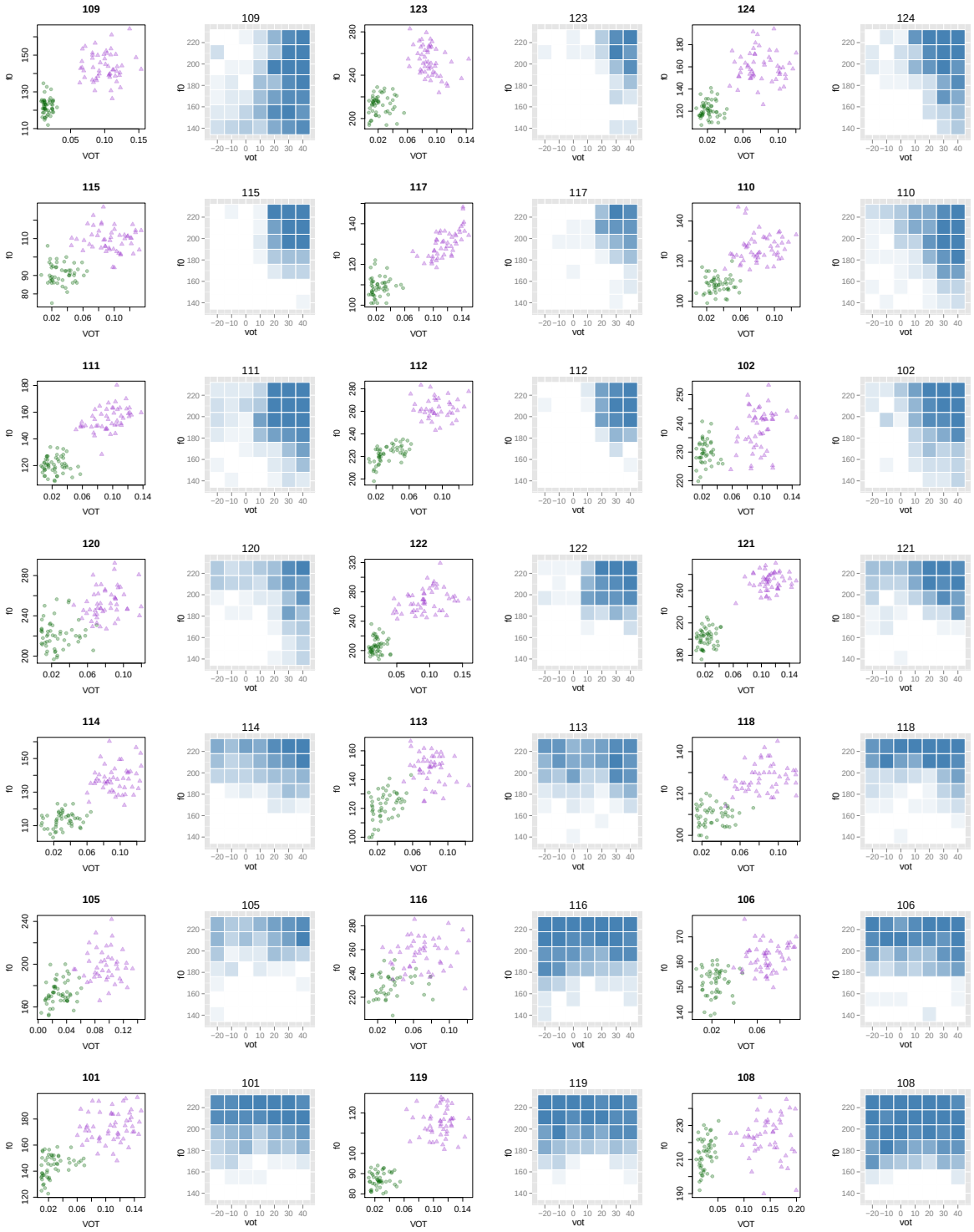



Figure 11: Plots of each subject's production results, with corresponding perception plots to the right. For perception plots, cell darkness represents percent "voiceless" response: dark blue = 100% voiceless; white = 100% voiced. Individuals are ranked by relative use of VOT to f0 in perception (Section 3.2.3).

| | Voiced vs. Voiceless | | | Aspirated vs. Lenis | | | Fortis vs. Lenis | | | Aspirated vs. Fortis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sub. | VOT | f0 | clos. | VOT | f0 | clos. | VOT | f0 | clos. | VOT | f0 | clos. |
| 108 | **2.93** | 0.01 | 0.35 | 0.34 | **−3.42** | −0.04 | 0.95 | **−3.26** | −0.47 | **−1.40** | −1.16 | 0.76 |
| 102 | **2.93** | 0.08 | 0.11 | 0.88 | **−2.89** | 0.30 | 1.14 | **−2.95** | 0.53 | **−2.04** | 0.71 | −0.42 |
| 106 | **2.22** | 0.49 | 1.03 | −0.67 | **−2.12** | 0.11 | 0.90 | **−1.72** | 0.23 | **−1.79** | 0.07 | 0.13 |
| 120 | **1.65** | 0.46 | 0.72 | 1.15 | **−2.14** | −0.21 | 1.11 | **−1.66** | −0.11 | **−2.17** | 0.19 | −0.62 |
| 109 | **2.96** | 0.91 | −0.15 | −1.81 | **−3.15** | 0.22 | 1.51 | **−2.44** | −0.15 | **−2.89** | −1.06 | 0.66 |
| 105 | **1.59** | 0.58 | 0.20 | −1.68 | **−2.74** | 0.19 | 2.02 | **−2.58** | 0.04 | **−2.23** | −0.88 | 0.30 |
| 117 | **2.18** | 0.96 | 0.21 | −0.38 | **−3.65** | −0.25 | 1.63 | **−2.21** | 0.22 | **−1.72** | −1.70 | −0.17 |
| 101 | **1.74** | 0.85 | 0.38 | −1.98 | **−2.34** | 0.19 | **1.88** | −1.15 | 0.11 | **−2.24** | −0.19 | 0.19 |
| 123 | **1.78** | 0.88 | 0.75 | 0.76 | **−3.06** | 0.28 | 1.60 | **−3.81** | 0.31 | **−1.69** | −0.93 | 0.23 |
| 113 | **1.45** | 0.78 | 0.22 | 0.55 | **−1.88** | −0.60 | **1.26** | −1.07 | −0.82 | **−1.73** | −0.19 | 0.40 |
| 124 | **2.05** | 1.22 | 0.36 | −0.75 | **−2.54** | −0.15 | **2.09** | −1.75 | −0.11 | **−2.61** | −0.30 | 0.19 |
| 110 | **1.51** | 1.11 | 0.65 | 0.07 | **−2.31** | −0.07 | 1.21 | **−1.87** | −0.43 | **−1.76** | −0.80 | 0.24 |
| 116 | 0.79 | 0.64 | **1.12** | −0.25 | **−2.97** | 0.13 | **1.67** | −1.57 | 0.12 | **−2.42** | −0.49 | 0.21 |
| 115 | **1.30** | **1.30** | 0.13 | 0.16 | **−3.10** | 0.39 | **1.94** | −1.88 | −0.06 | **−2.27** | −1.18 | 0.41 |
| 119 | 1.86 | **1.87** | 0.21 | −2.00 | **−3.01** | −0.19 | 1.65 | **−1.70** | 0.25 | **−2.44** | −0.81 | −0.07 |
| 122 | 1.54 | **1.60** | 0.13 | 0.95 | **−2.39** | −0.73 | **2.76** | −0.82 | −0.30 | **−2.88** | 0.08 | −0.05 |
| 114 | 1.20 | **1.32** | 0.40 | −0.31 | **−4.13** | 0.23 | 1.83 | **−5.05** | 0.54 | **−2.76** | −0.89 | −0.11 |
| 118 | 1.06 | **1.29** | −0.01 | −1.73 | **−1.95** | 0.28 | **1.50** | −1.32 | 0.07 | **−1.74** | −0.25 | 0.06 |
| 121 | 1.54 | **2.40** | 0.08 | −0.21 | **−3.05** | −0.07 | **2.03** | −1.50 | −0.42 | **−2.33** | −1.39 | 0.07 |
| 112 | 0.89 | **1.80** | 0.59 | −0.07 | **−4.37** | −0.01 | **3.10** | −2.53 | 0.29 | **−3.82** | 0.04 | −0.46 |
| 111 | 0.88 | **2.08** | −0.18 | −0.19 | **−2.60** | 0.00 | **2.26** | −1.39 | 0.04 | **−2.35** | −0.63 | −0.06 |
| **Mean** | **1.72** | 1.08 | 0.35 | −0.34 | **−2.85** | 0.00 | 1.72 | **−2.11** | −0.01 | **−2.25** | −0.56 | 0.09 |

Table 12: Individual and group coefficients from discriminant analyses in English and in each pairwise comparison of the Korean three-way contrast. The first stop type listed in each comparison is the "default" type; the direction of the coefficients represents the direction of that dimension that best predicts the second stop type (e.g. the positive coefficients for VOT in the English contrast represent the model's use of higher values of VOT to predict voiceless stops, the non-default category). Participants are sorted by the relative reliability of VOT as compared to f0 in distinguishing each speaker's production of each contrast (with the absolute values of the two coefficients adding to 1); the most reliable cue is indicated in bold. Individual subject numbers are included, here and throughout the paper, for purposes of comparison across experiments.

| | Voiced vs. Voiceless | | | Aspirated vs. Lenis | | | Fortis vs. Lenis | | | Aspirated vs. Fortis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sub. | VOT | f0 | clos. | VOT | f0 | clos. | VOT | f0 | clos. | VOT | f0 | clos. |
| 109 | **2.02** | −0.39 | −0.08 | −0.93 | **−3.72** | −0.50 | 2.32 | **−2.59** | 0.10 | **−3.71** | −0.13 | −0.34 |
| 123 | **2.88** | 1.76 | 1.14 | −1.40 | **−3.87** | −0.91 | 2.49 | **−2.72** | −1.31 | **−7.05** | −1.06 | 0.21 |
| 124 | **2.61** | 1.79 | 0.63 | −1.75 | **−3.72** | −0.06 | 2.12 | **−3.36** | −0.58 | **−4.13** | −0.30 | 0.16 |
| 115 | **3.46** | 2.57 | −0.13 | −1.64 | **−4.60** | −0.34 | 1.92 | **−4.75** | −0.40 | **−7.84** | −0.51 | −0.02 |
| 117 | **2.75** | 2.04 | 0.81 | −2.62 | **−4.88** | −0.36 | 2.37 | **−3.09** | −0.54 | **−6.66** | −1.03 | 0.03 |
| 110 | **1.92** | 1.51 | 0.33 | −1.56 | **−5.79** | −0.66 | 2.52 | **−5.71** | −0.09 | **−4.26** | −0.18 | 0.05 |
| 111 | **1.88** | 1.48 | 0.35 | −1.28 | **−4.04** | −0.44 | 4.86 | **−6.02** | −0.79 | **−6.03** | 0.66 | −0.22 |
| 112 | **3.54** | 2.77 | 0.07 | −1.77 | **−5.07** | −0.66 | 1.90 | **−4.72** | −0.62 | **−5.66** | −0.05 | 0.14 |
| 102 | **2.16** | 1.76 | 0.63 | −1.07 | **−5.21** | −0.78 | 3.09 | **−4.58** | −1.08 | **−6.14** | 1.04 | 0.01 |
| 120 | 1.87 | **1.90** | 1.74 | −1.81 | **−4.84** | −0.93 | 4.91 | **−5.35** | −1.16 | **−9.83** | −1.26 | −0.28 |
| 122 | 2.18 | **2.43** | 0.38 | −2.76 | **−5.16** | −0.71 | 3.08 | **−5.52** | −1.33 | **−6.78** | −0.88 | 0.00 |
| 121 | 1.30 | **2.73** | 1.01 | −2.42 | **−7.29** | −0.64 | 1.67 | **−3.12** | −0.41 | **−6.62** | −2.36 | 0.20 |
| 114 | 0.89 | **2.44** | 0.48 | −1.29 | **−5.76** | −1.03 | 3.72 | **−4.66** | −0.88 | **−5.90** | −0.27 | 0.36 |
| 113 | 0.43 | **2.09** | 0.30 | −1.40 | **−3.02** | 0.02 | 2.13 | **−3.50** | −0.15 | **−4.70** | 0.57 | −0.19 |
| 118 | 0.59 | **3.05** | 0.55 | −2.13 | **−7.26** | −0.94 | 2.69 | **−6.97** | −0.39 | **−6.62** | −0.34 | −0.07 |
| 105 | 0.37 | **2.08** | 0.26 | −1.40 | **−4.91** | −0.58 | 3.93 | **−6.10** | −1.07 | **−3.73** | 0.13 | 0.00 |
| 116 | −0.49 | **3.11** | 0.55 | −1.64 | **−9.06** | −0.13 | 3.80 | **−7.10** | −0.09 | **−7.50** | 1.16 | 0.00 |
| 106 | 0.38 | **2.93** | 0.50 | −1.75 | **−10.12** | −0.81 | 1.53 | **−9.31** | −0.23 | **−4.26** | 0.27 | −0.35 |
| 101 | 0.17 | **2.83** | 0.33 | −2.51 | **−7.77** | −0.23 | 1.41 | **−4.69** | −0.09 | **−4.19** | −0.31 | 0.11 |
| 119 | −0.21 | **3.35** | 0.22 | −1.35 | **−4.85** | −0.62 | **3.55** | −2.70 | −0.19 | **−7.34** | −1.00 | −0.30 |
| 108 | −0.19 | **3.79** | 0.44 | −0.80 | **−5.07** | −0.83 | 3.02 | **−5.45** | −0.78 | **−4.29** | 0.30 | 0.17 |
| **Mean** | 1.45 | **2.29** | 0.50 | −1.68 | **−5.52** | −0.58 | 2.81 | **−4.86** | −0.58 | **−5.87** | −0.26 | −0.02 |

Table 13: Beta-coefficients from binary logistic regression models examining how variation in each acoustic parameter predicts individual listener responses for the English stop voicing contrast and each pairwise Korean contrast. The first stop type listed in each comparison is the "default" type; coefficients represent increased (positive coefficients) or decreased (negative coefficients) log odds of the other member of the contrast when the given acoustic parameter increases by one standard deviation. Participants are sorted by the relative predictability of VOT as compared to f0 in predicting response (with the absolute values of the two coefficients adding to 1).

References

Ashby, F. G., Queller, S., and Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, 61(6):1178–1199.

Baker, W. and Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, 48(1):1–27.

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

Best, C. T. and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.-S. and Munro, M. J., editors, *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, Amsterdam. John Benjamins.

Boersma, P. and Weenink, D. (2011). Praat: doing Phonetics by computer, version 5.3: http://www.praat.org.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In Strange, W., editor, *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research*, pages 275–300. York Press, Timonium, MD.

Bohn, O.-S. and Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11:303–328.

Bohn, O.-S. and Flege, J. E. (1997). Perception and production of a new vowel category by second-language learners. In James, A. and Leather, J., editors, *Second-language speech: Structure and process*, pages 53–74. Walter de Gruyter, Berlin.

Bradlow, A., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4):2299–2310.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3):372–387.

Chandrasekaran, B., Sampath, P., and Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, 128:456–465.

Cho, T., Jun, S.-A., and Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30(2):193–228.

Cho, T. and Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29(2):155–190.

Cho, T. and Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, 37(4):466–485.

Díaz, B., Baus, C., Escera, C., Costa, A., and Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences*, 105(42):16083–16088.

Dmitrieva, O., Llanos, F., Shultz, A. A., and Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49:77–95.

Escudero, P., Benders, T., and Lipski, S. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, 37(4):452–465.

Escudero, P. and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26:551–585.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, pages 233–277. York Press, Timonium, MD.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In Cole, J. and Hualde, J., editors, *Laboratory Phonology 9*, pages 353–380. Mouton de Gruyter, Berlin.

Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25:437–470.

Flege, J. E. and MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26(1):1–34.

Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *The Journal of the Acoustical Society of America*, 106:2973–2987.

Flege, J. E., Takagi, N., and Mann, V. A. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*, 99(2):1161–1173.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist approach. *Journal of Phonetics*, 14:3–28.

Francis, A., Kaganovich, N., and Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, 124:1234.

Goudbeek, M., Cutler, A., and Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, 50(2):109–125.

Green, K. P., Zampini, M. L., and Clarke, C. M. (1998). The role of preceding closure interval and voice onset time in the perception of voicing: a comparison of English versus Spanish-English bilinguals. *The Journal of the Acoustical Society of America*, 104(3):1835.

Hattori, K. and Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, 125(1):469–479.

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, 108(6):3013–3022.

Holt, L. and Lotto, A. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119:3059–3071.

Idemaru, K., Holt, L., and Seltman, H. (2012). Individual differences in cue weights are stable across time: the case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, 132(6):3950–3964.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1):B47–B57.

Kang, K.-H. (2009). *Clear speech production and perception of Korean stops and the sound change in Korean stops*. PhD thesis, University of Oregon.

Kang, K.-H. and Guion, S. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *The Journal of the Acoustical Society of America*, 119(3):1672–1683.

Kang, Y. (2008). Tensification of voiced stops in English loanwords in Korean. *Harvard studies in Korean Linguistics*, 12:179–192.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45:76–90.

Kim, M.-R. C. (1994). *Acoustic characteristics of Korean stops and perception of English stop consonants*. PhD thesis, University of Wisconsin-Madison.

Kim, S. and Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *The Journal of the Acoustical Society of America*, 134(1):EL19–EL25.

Kingston, J., Diehl, R., Kirk, C., and Castleman, W. (2008). On the initial perceptual structure of distinctive features: the [voice] contrast. *Journal of Phonetics*, 36:28–54.

Kondaurova, M. and Francis, A. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, 124(6):3959–3971.

Kong, E. and Yoon, I. (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.

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

Lee, H., Politzer-Ahles, S., and Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, 41(2):117–132.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21:1–36.

Lisker, L. and Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20:384–422.

Llanos, F., Dmitrieva, O., Shultz, A., and Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, 134(3):2213–2224.

Löfqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, 85(3):1314–1321.

Lotto, A., Sato, M., and Diehl, R. (2004). Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/. In Slifka, J., Manuel, S., and Matthies, M., editors, *From Sound to Sense: 50+ Years of Discoveries in Speech Communication*, pages 381–386.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Attention, Perception, & Psychophysics*, 18(5):331–340.

Morrison, G. (2009). L1-Spanish speakers acquisition of the English/i/-/i/contrast II: Perception of vowel inherent spectral change. *Language and Speech*, 52(4):437–462.

Morrison, G. S. (2005). An appropriate metric for cue weighting in L2 speech perception. *Studies in Second Language Acquisition*, 27:597–606.

Morrison, G. S. (2007). Logistic regression modelling for first- and second-language perception data. In Solé, M. J., Prieto, P., and Mascaró, J., editors, *Segmental and prosodic issues in Romance phonology*, pages 219–2346. John Benjamins, Amsterdam.

Morrison, G. S. (2008). L1 Spanish speakers' acquisition of the English /i/-/i/ contrast: Duration-based perception is not the initial developmental stage. *Language and Speech*, 51(4):285–315.

Morrison, G. S. and Kondaurova, M. V. (2009). Analysis of categorical response data: Use logistic regression rather than endpoint-diferent scores or discriminant analysis. *The Journal of the Acoustical Society of America*, 126(5):2159–2162.

Nearey, T. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, 101(6):3241–3254.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, 113(5):2850–2857.

Park, H. and de Jong, K. (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.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., and Stockmann, E. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4):2338–2344.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, 89(6):2961–2977.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In Strange, W., editor, *Speech perception and linguistic experience: Issues in cross-language research*, pages 379–410. York Press.

Schertz, J., Cho, T., Lotto, A., and Warner, N. (submitted). Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics*.

Schmidt, A. M. (1996). Cross-language identification of consonants, part 1. Korean perception of English. *The Journal of the Acoustical Society of America*, 99(5):3201–3211.

Sebastián-Gallés, N. and Baus, C. (2005). On the relationship between perception and production in L2 categories. In Cutler, A., editor, *Twenty-first century psycholinguistics: Four cornerstones*, pages 279–292. Erlbaum, New York.

Shultz, A., Francis, A., and Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2):EL95–EL101.

Silva, D. (1992). *The phonetics and phonology of stop lenition in Korean*. PhD thesis, Cornell University.

Silva, D. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23(2):287–308.

Stevens, K. N. (1959). Effect of duration upon vowel identification. *The Journal of the Acoustical Society of America*, 31:109.

Toscano, J. C. and McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 34(3):434–464.

Wang, Y., Jongman, A., and Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113:1033.

Wanrooij, K., Escudero, P., and Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, 41(5):307–319.