Journal of Phonetics 50 (2015) 1–14

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

A longitudinal study of the second language acquisition of a three-way stop contrast

Jeffrey J. Holliday[a,b,*]

[a] Department of Second Language Studies, Indiana University, Memorial Hall, 1021 E. Third St., Bloomington, IN 47405, USA
[b] Department of Linguistics, The Ohio State University, Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 8 October 2014
Received in revised form
18 January 2015
Accepted 26 January 2015

Keywords:
Korean
Mandarin
Stop consonants
Second language acquisition

ABSTRACT

The goal of this paper was to document how native (L1) speakers of a language with a two-way stop contrast acquire a three-way stop contrast in a second language (L2). Mandarin presents a two-way stop contrast cued primarily by VOT, whereas Korean presents a three-way stop contrast cued jointly by VOT and the *f*0 of the following vowel. Mandarin and Korean stop productions from 12 L1 Mandarin novice L2 learners of Korean were subjected to acoustic analysis. Results revealed a wide variety of production patterns, suggesting that the learning of an L2 contrast may not always be predicted by cross-language acoustic correspondences. Six of the participants were recorded again both six and 12 months later. The longitudinal results showed that some learners were unable to produce the Korean contrast in a native-like way even after one year of intensive L2 instruction. Learners whose initial production strategy was consistent but incorrect fared worse after one year than learners whose productions initially exhibited more variability. These results contribute to our understanding of both the L2 acquisition of “new” and “similar” categories and also how well naïve perceptual assimilation can predict L2 production.

1. Introduction

Stop consonants have been widely investigated in the L2 literature, with most studies focusing on the L2 acquisition of a two-way stop contrast by L1 speakers of languages that also have a two-way stop contrast (e.g., Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Flege, 1991; Flege & Eefting, 1987, 1988). In the vast majority of these studies, the stop contrasts in both the L1 and the L2 are differentiated primarily by voice onset time (VOT), which is a temporal measure that relates the stop burst and an onset of voicing. When the onset of voicing precedes the stop burst the stop is said to be characterized by lead VOT (or “pre-voicing”), and when the stop burst precedes the onset of voicing it is described as having lag VOT. Lisker and Abramson (1964) showed that cross-linguistic differences in stop contrasts are often manifested through how VOT is distributed across each language's stop categories. For example, some languages, like Dutch, exhibit a two-way stop contrast in word-initial position between stops with lead VOT and short lag VOT, whereas other languages, like English, contrast stops with short and long lag VOT. Although this English stop contrast is generally thought of as one of voicing rather than aspiration, in word-initial position the “voiced” stops are typically realized with short lag VOT (Ladefoged, 1999).

When two languages use VOT to cue stop contrasts differently, L1 speakers of one of the languages learning the other language as an L2 will often fail to produce L2 stops with native-like VOT. For example, when L1 speakers of a language that has voiceless stops with short lag VOT (represented as /p, t, k/), such as French, produce English voiceless stops, which have long lag VOT (and are also represented as /p, t, k/), the L2 English voiceless stops are typically realized with short lag VOT (Caramazza et al., 1973; Swanson, 2006), suggesting that the English voiceless aspirated stop categories have been phonologically equated with L1 voiceless unaspirated stop categories. Cases of VOT mismatches such as these are well documented in the literature, such as English and Dutch (Flege & Eefting, 1987), English and Spanish (Flege, 1991; Flege & Eefting, 1988), English and Brazilian

* Correspondence address: Department of Second Language Studies, Indiana University, Memorial Hall, 1021 E. Third St., Bloomington, IN 47405, USA. Tel.: +1 812 855 0032; fax: +1 812 855 5605.
E-mail address: jjhollid@indiana.edu

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2015.01.004

Portuguese (Sancier & Fowler, 1997), and Spanish and Swedish (Stölten, Abrahamsson, & Hyltenstam, in press). With increased L2 experience, L2 learners often learn to either differentiate between similar L1 and L2 categories or to produce the stops in both languages with a "compromise" VOT (Williams, 1980), such that, for example, more experienced L1 French learners of English would produce voiceless stops with lag VOT in French longer than monolingual French production and lag VOT in English shorter than monolingual English production, and vice versa for more experienced L1 English learners of French (e.g. Fowler, Sramko, Ostry, Rowland, & Hallé, 2008; Hazan & Boulakia, 1993). These findings are generally in line with the predictions made by the Speech Learning Model (SLM; Flege, 1995, p. 238), which proposes that L2 production is guided primarily by perceptual targets. According to this framework, for an L1 French learner of English, English /t/ would be produced with shorter lag VOT because it is "similar" to French /t/ (that is, perceptually, French /t/ is a similar enough French sound to English /t/), and an L1 French learner's perceptual target for /t/ is in the short lag VOT range. The SLM predicts that "new" categories, which are not perceived by the learner as being similar to an existing L1 category, may be easier to learn because they are not classified as equivalent to any native sound.

Accordingly, when both the L1 and L2 have a two-way stop contrast that can be differentiated by VOT alone, it is straightforward to predict which L2 category will be phonetically realized as which L1 category by novice L2 learners: the stop categories with the greater positive VOT values in both languages will be equated with each other, and then likewise with the stop categories in both languages with the lesser positive (or greater negative) VOT values. Crucially, such patterns found in native speakers and L2 learners of these European languages are reinforced by the orthographic equivalence between the categories (e.g. long lag English /t/ and short lag Dutch /t/ are both written as <t>). We should not expect cases in which, for example, an L1 Dutch speaker consistently produced English /t/ with pre-voicing and English /d/ with short lag VOT, presumably assimilating English /t/ to Dutch /d/ and English /d/ to Dutch /t/.

Stop contrasts in many languages, however, involve more than two categories, rely on acoustic cues other than VOT, or are not written with the same graphemes as the stops in well-studied European languages. The current study investigated the case of L1 speakers of Mandarin, a language with a two-way contrast between short-lag and long-lag VOT stops, learning Korean, a language with a three-way stop contrast differentiated by both VOT and *f*0 in word-initial position. In this scenario, the L2 learner cannot simply substitute a different L1 stop category for each L2 stop category in the same way that an L1 Dutch speaker could substitute Dutch /d/ for English /d/ and Dutch /t/ for English /t/. If an L2 learner of Korean tried this, he or she would necessarily be forced to use the same L1 category for two different L2 categories: while such a strategy is certainly possible, it would not result in production that would be perceived as correct by native Korean listeners.

1.1. L2 acquisition of Korean stops

Seoul Korean (hereafter, Korean) has a three-way phonation type stop contrast between fortis (tense), lenis (lax), and aspirated stops, which have historically been described as having short lag, intermediate lag, and long lag VOT, respectively, in word-initial position. VOT was supported by a secondary cue, the *f*0 of the following vowel, which helped to distinguish lenis stops, with a low *f*0, from fortis and aspirated stops, which had a high *f*0 (Han & Weitzman, 1970; Kagaya, 1974). More recent acoustic studies of younger Korean speakers have continued to find the same *f*0 difference between lenis stops and fortis and aspirated stops (cf. Kang, 2014; Kang & Guion, 2008; Silva, 2006), but have also found that this younger generation of speakers has neutralized the difference in VOT between the lenis and aspirated stops in word-initial position, with both being realized with long lag VOT. This sound change has resulted in the word-initial lenis-aspirated contrast being realized as an *f*0 difference, whereas VOT still functions to differentiate fortis stops from both lenis and aspirated stops.

Studies of non-native perception of the Korean stop contrast have shown that fortis stops are more readily perceived as different from lenis and aspirated stops, and that lenis and aspirated stops are often perceived as similar to each other. In perceptual assimilation studies that have tested L1 listeners from languages with a two-way VOT stop contrast, such as Japanese (Yasuta, 2004), English (Schmidt, 2007), and Mandarin (Holliday, 2014), it was found that Korean fortis stops were perceived as unaspirated stops (or voiced or short-lag VOT) in the vast majority of trials, and both Korean lenis and aspirated stops were perceived as aspirated (or voiceless or long-lag VOT), with lenis stops sometimes being a slightly less good category fit depending on the place of articulation and vowel context.

Thus, it has been shown that Korean lenis and aspirated stops are often perceived as the same category, and not the same category as fortis stops, by non-native listeners whose L1 has a two-way VOT stop contrast. The Perceptual Assimilation Model (PAM; Best, 1995), which makes predictions about the relationship between perceptual assimilation and discrimination of non-native speech sounds by naïve listeners, predicts that discrimination between non-native sounds that assimilate to a single L1 category should be poor, and discrimination between sounds that assimilate to different L1 categories should be good. PAM therefore predicts that discrimination between Korean lenis and aspirated stops by naïve L1 Mandarin listeners should be poor, and discrimination between Korean fortis and either lenis or aspirated stops should be good.

The perception of Korean stops by L2 learners, on the other hand, could be quite different. In the extension of PAM to L2 learners, PAM-L2, Best and Tyler (2007) point out that the linguistic experience of L2 learners differs from naïve listeners in ways that could affect how they assimilate L2 sounds to L1 categories. For example, the phonological status or orthographic representation of an L2 sound is both irrelevant and inaccessible to a naïve listener, but such information could be very salient to an L2 learner and could certainly impact perception. Best and Tyler (2007, p. 28) cite the example of French /r/ (realized as [ʁ]), which although is acoustically quite distinct from English /r/ (realized as [ɹ]), could nonetheless be assimilated to English /r/ at the phonological level. Thus, because the L2 learner must process both phonetic and phonological information, L2 learners' perception could diverge from that of naïve listeners from the same L1 background, who can only process the phonetic information. Following SLM (Flege, 1995), L2 learners' productions could subsequently diverge from what would be predicted by the perception of naïve listeners.

Turning back to Korean stops, Jung and Kwon (2010) found that L2 Korean learners from a variety of L1 backgrounds were much less accurate at discriminating between lenis and aspirated stops (25.3%, with a 20% chance level), than between lenis and fortis stops (81.3%) or fortis and aspirated stops (76.3%). As these results are what PAM (Best, 1995) would predict based on the perceptual assimilation patterns of naïve listeners, they are also predicted by PAM-L2, but only if the L2 learners were processing the stimuli at the phonetic level or if L2 phonological information did not interfere with their perception.

Based on the results of Jung and Kwon (2010), and following SLM (Flege, 1995), we should expect L1 Japanese, English, or Mandarin learners of Korean to produce more of a difference between fortis stops versus lenis and aspirated stops, and less of a difference between lenis and aspirated stops themselves. In a longitudinal production study, Chang (2010) tracked the use of VOT and *f*0 in stop productions by 26 L1 English learners of L2 Korean over the course of a six-week intensive language course. Although the participants all shared an L1 background and none had had any exposure to Korean prior to beginning the language course, the subjects exhibited a wide variety of production patterns, many of which could not be predicted by how naïve L1 English listeners (cf. Schmidt, 2007) or L2 Korean learners (Jung & Kwon, 2010) perceive Korean stops. For example, seven of the speakers produced both lenis and fortis stops with short lag VOT and aspirated stops with long lag VOT, ostensibly substituting English voiced /b, d, g/ for both Korean lenis /p, t, k/ and fortis /p*, t*, k*/, and English voiceless /p, t, k/ for Korean aspirated /p^h, t^h, k^h/. These seven speakers also used *f*0 in different ways, with some speakers producing lenis stops with low *f*0 and fortis stops and aspirated stops with high *f*0, and other speakers producing fortis stops with low *f*0 and lenis stops with high *f*0. Other learners made a three-way contrast, but not necessarily in a native-like way, and yet other learners produced no clear contrast among any of the categories at all.

Chang (2010) offered several possible explanations for his findings, including the influences of orthography, classroom input, individual differences in L1 cue weighting, and the use of explicit strategies that "may or may not be based on actual L2 input patterns" (Chang, 2010, p. 102). These results, however, raise two additional sets of questions. First, does an L2 learner's particular use of acoustic cues (i.e. weighting of VOT and *f*0) at the novice stage determine the trajectory of cue use at later stages of acquisition? Specifically, among subjects whose production was markedly non-native-like, did some have a more positive prognosis than others? In a similar vein, studies of covert contrast in children have shown not only that there are different ways to incorrectly produce a phonological contrast (e.g. Li, Edwards, & Beckman, 2009; Macken & Barton, 1980) but that children's later phonological development may be related to these error patterns at earlier stages of acquisition (Forrest, Dinnsen, & Elbert, 1997; Tyler, Figurski, & Langsdale, 1993). Because most longitudinal studies of adult L2 production involve learners who have already had several years of exposure to the target language at the beginning of the study (Aoyama, Guion, Flege, Yamada, & Akahane-Yamada, 2008; Oh et al., 2011; Tsukada et al., 2004) more longitudinal studies of novice L2 learners are needed before we can understand the predictive power of novice error patterns on later L2 production.

Second, to what extent was the wide variety of production strategies found in Chang (2010) L1-specific? Although Mandarin stops are cued primarily by VOT, much like English, would L1 speakers of Mandarin, a tonal language, learn to manipulate the *f*0 cue earlier and more accurately than L1 speakers of English, a non-tonal language? Previous studies have shown that native speakers of tonal languages are often more accurate than non-tonal language speakers on some tone perception tasks (e.g. So & Best, 2010; Wayland & Guion, 2004), although native experience with a tone language does not automatically translate into across-the-board higher accuracy in non-native tone perception (Francis, Ciocca, Ma, & Fenn, 2008), or on non-linguistic pitch perception tasks (Alexander, Bradlow, Ashley, & Wong, 2011). Furthermore, L1 speakers of non-tonal languages are still able to perceive some differences in non-native tonal contrasts (Hallé, Chang, & Best, 2004), so it is not clear that L1 Mandarin learners of Korean would have an outright advantage over L1 English learners in acquiring the *f*0 cue to the Korean stop contrast.

The current study explored these issues through a longitudinal investigation of the L2 acquisition of the Korean stop contrast by L1 speakers of Mandarin. Whereas Chang (2010) looked at only the first 6 weeks of L2 learning, and was thus able to track changes at the absolute earliest stage of acquisition, the current study began after 6 weeks of learning and was repeated six and twelve months later to capture a longer-term trajectory of learning. Our two main questions were: (1) how did the L2 learners use VOT and *f*0 to cue the three-way contrast, and (2) did the learners produce the Korean stops in the same way as their native Mandarin stops? It was predicted that L1 Mandarin learners of Korean would initially produce Korean fortis stops as exemplars of Mandarin unaspirated stops, and Korean lenis and aspirated stops as exemplars of Mandarin aspirated stops, following the ostensibly VOT-guided naïve perceptual assimilation patterns documented in Holliday (2014). It was also predicted that the learners would successfully manipulate the *f*0 cue to differentiate Korean lenis and aspirated stops, as the *f*0 difference between the two stop categories should be much more salient to L1 Mandarin speakers than to the L1 English speakers tested in Chang (2010).

2. Experiment 1

2.1. Methods

2.1.1. Participants

The participants were 12 L1 Mandarin learners of Korean (11 female, 1 male) who had been studying Korean at the absolute beginner level for 6 weeks at a university in Seoul. Participants were in Korean class four hours per day, 5 days per week. At this stage in their program, students had learned how to read, use basic greetings, and form simple sentences. All participants were born and raised in different provinces across mainland China. Their ages ranged from 19 to 27 years old, with the mean and median age at 21.5 years. None reported having studied any foreign language other than Korean or English, and all but two had arrived in Korea the day before their language program began. Two subjects, c23 and c24, had been in Korea for approximately 4.5 months. The subjects were not all in the same class, but they were being taught in the same program by teachers who were following a common curriculum and using the same textbook and materials.

2.1.2. Materials

Participants first read aloud a Korean word list containing 51 stop-initial target words (17 per phonation type) mixed with fillers for a total of 132 words. Each target word began with an alveolar or velar stop followed by one of the vowels /a/, /i/, /u/, or /ɨ/. There were no Korean alveolar target words in the /i/ context because most Korean words beginning with those CVs are loanwords from English that may have been too unfamiliar to novice L1 Mandarin learners. Participants also read aloud a Mandarin word list containing 18 stop-initial target words (9 per stop category) mixed with fillers for a total of 131 words. Each Mandarin target word began with an alveolar stop followed by one of the vowels /a/, /i/, or /u/. Thus, in total, each subject produced 69 stop-initial CVs, the distribution of which is summarized in Table 1.

2.1.3. Procedure

Participants read the Korean word list first and the Mandarin word list second. They were seated at a desk or table in front of a University Sounds US658H microphone mounted on a table-top stand approximately 15–20 cm from the participant's mouth. Participants were instructed to read the list at a comfortable pace, inserting a pause of at least 1 s between words, and to avoid using list intonation. The recordings were done with the microphone connected either to a laptop computer through a Roland UA-30 USB interface or directly to a Roland Edirol digital recorder. All recordings were sampled at 44.1 kHz.

2.1.4. Data analysis

VOT was measured in each token as the time between the peak energy of the stop burst and the onset of voicing, measured as the first upswing zero crossing at the beginning of periodicity in the waveform. *f*0 was measured as the mean over the first 15 ms of the vowel, using the autocorrelation method in Praat (Boersma & Weenink, 2013). *f*0 was not measured at vowel onset because some tokens either exhibited rapid *f*0 fluctuation or did not have a measurable *f*0 at vowel onset, and previous studies of native Korean stop production that have measured *f*0 as a mean over an even longer interval (Holliday & Kong, 2011) or at the vowel midpoint (Kang & Guion, 2006) have still shown robust *f*0 differences among phonation types. Finally, *f*0 values for the Korean target productions were *z*-score normalized within each subject, and for those who participated in the longitudinal portion of the study, *f*0 was normalized across all of the subject's Korean target productions across all time points. The subjects' Mandarin target productions were not included in the normalization as the potential for overall *f*0 differences in language mode (Altenberg & Ferrand, 2006; Ng, Chen, & Chan, 2012) could obfuscate the interpretation of subjects' use of *f*0 in Korean.

2.2. Results

2.2.1. Production of Korean alveolar stops

Our first research question was whether the subjects distinguished between Korean stop categories in terms of VOT, *f*0, both, or neither. The alveolar stops were first examined by running a MANOVA with dependent variables of VOT and *f*0 and independent variables of phonation type (fortis, lenis, aspirated) and subject. There was a statistically significant overall effect of phonation type, and individual ANOVAs for both VOT and *f*0 revealed a significant relationship with phonation type for VOT but not *f*0. The strength of these relationships varied across subjects, however, as the interaction between phonation type and subject was significant for both VOT and *f*0.

The mean VOT for each phonation type was 31.4 ms for /t*/, 41.2 ms for /t/, and 62.0 ms for /tʰ/. Post-hoc Tukey HSD tests showed that across all subjects the differences between /t/-/t*/ ($p=.017$), /t/-/tʰ/ ($p<.001$), and /t*/-/tʰ/ ($p<.001$) were statistically significant. The mean normalized *f*0 for each phonation type was .054 for /t*/, −.076 for /t/, and .168 for /tʰ/. Although these differences fall in the Korean native-like order of lenis < fortis < aspirated, post-hoc Tukey HSD tests showed that none of these differences were statistically significant. However, even though the post-hoc tests took between-subject variation into account, the individual distributions of VOT and *f*0 shown in Fig. 1 suggest that very few individual subjects exhibited a robust difference among all three phonation types. That is, while many of the subjects' VOT distributions exhibited a trend of fortis < lenis < aspirated, there was

Table 1
Number of target words for each stop × vowel combination.

Stop category	Vowel context			
	/a/	/i/	/u/	/ɨ/
Korean /t*/, /t/, /tʰ/	3	0	3	1
Korean /k*/, /k/, /kʰ/	3	3	3	1
Mandarin /t/, /tʰ/	3	3	3	0

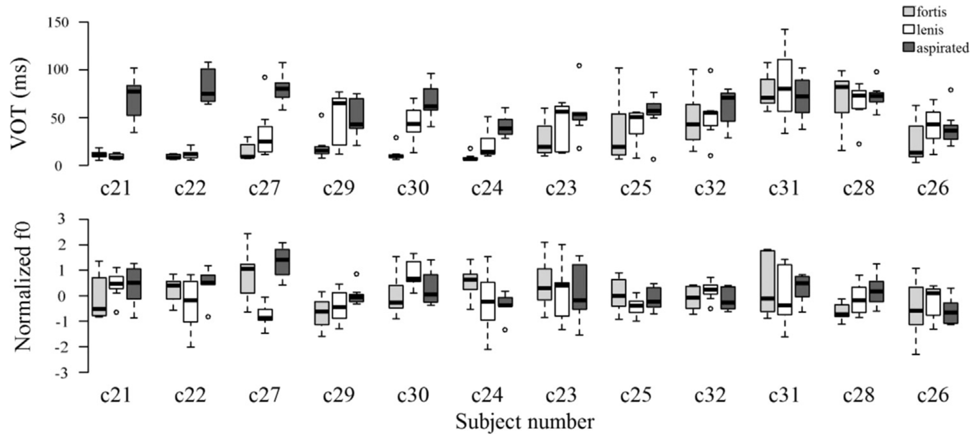

Fig. 1. Boxplots of VOT and *f*0 for each Korean alveolar phonation type for each subject. The subjects are ordered according to the order in which they are discussed.

Table 2
MANOVA and individual ANOVA output for alveolar and velar Korean stop productions.

	MANOVA (Wilks' lambda)	ANOVA	Main effect	Subject interaction
Alveolar	$F(4,430)=19.17^{***}$	VOT	$F(2,216)=38.8^{***}$	$F(22,216)=3.38^{***}$
		*f*0	$F(2,216)=2.03^{n.s.}$	$F(22,216)=2.23^{**}$
Velar	$F(4,646)=37.06^{***}$	VOT	$F(2,324)=73.74^{***}$	$F(22,324)=4.18^{***}$
		*f*0	$F(2,324)=5.87^{***}$	$F(22,324)=1.20^{n.s.}$

Note: $^{*}p<.05$.
$^{**}\ p<.01$.
$^{***}\ p<.001$.

substantial inter-subject variation in terms of whether individual categories were realized as short or long lag VOT, and very few subjects seemed to clearly differentiate among all three phonation types.

Specifically, some subjects, such as the first four from the left in Fig. 1 (c21–c29), exhibited what appears to be a two-way VOT contrast, with c21, c22, and c27 distinguishing short lag VOT for /t*/ and /t/ and long lag VOT for /tʰ/, and subject c29 showing a somewhat less clear distinction between short-lag /t*/ and long-lag /t/ and /tʰ/. This pattern exhibited by subject c29 lines up most closely with how L1 Korean speakers produce the stop contrast (cf. Holliday & Kong, 2011, for native Korean data elicited from the same experiment), although the wide variation in lenis stop VOT indicates that subject c29 did not produce lenis stops with consistently long lag VOT. Other subjects exhibited a variety of patterns and a range of degrees of differentiation among phonation types. In the most extreme cases, subjects c31 and c28 seemed to produce all three phonation types with entirely overlapping VOT distributions.

Turning to *f*0, in the bottom panel of Fig. 1, it is immediately clear that subject c27 was the only one to successfully implement the *f*0 cue in a native-like way. Some subjects, such as c22, c25, and c31, exhibited the correct trend, with lenis stops having a lower *f*0 than fortis or aspirated stops, but because the distributions for each phonation type overlapped so much it is difficult to classify them as native-like. Overall, these results suggest that at this early stage of L2 learning the majority of subjects have not learned how to manipulate the *f*0 cue in a native-like way.

2.2.2. Production of Korean velar stops

To analyze the subjects' Korean velar stops, an analogous set of MANOVA and ANOVA analyses was run. The results, shown in to bottom half of Table 2, indicate a statistically significant overall effect of phonation type, which was found to be significant for both VOT and *f*0. Only the differences in VOT depended on subject, however, as the interaction between phonation type and subject was significant for VOT but not *f*0.

The mean VOT for each phonation type was 40.1 ms for /k*/, 62.3 ms for /k/, and 79.7 ms for /kʰ/. Post-hoc Tukey HSD tests showed that the pairwise differences among all phonation types were statistically significant ($p<.001$). As with the alveolar stops, it is clear from Fig. 2 that many individual subjects did not display this pattern, once again demonstrating that an overall significant trend may only describe some of the subjects' individual patterns. Subjects c21 and c22 produced a clear difference between fortis and lenis stops (with short-lag VOT) and aspirated stops (with long-lag VOT). Other subjects, such as c27, c29, c30, and c24, exhibited what may be a two-way VOT contrast, and the remaining subjects exhibited a range of patterns and degrees of overlap between phonation types.

For *f*0, the mean normalized *f*0 for each phonation type was .134 for /k*/, −.141 for /k/, and .297 for /kʰ/. Post-hoc Tukey tests showed that only the difference between /k/ and /kʰ/ was statistically significant ($p<.001$). The bottom panel of Fig. 2 shows that subjects did not differentiate among the three phonation types, with the exception of subject c27.

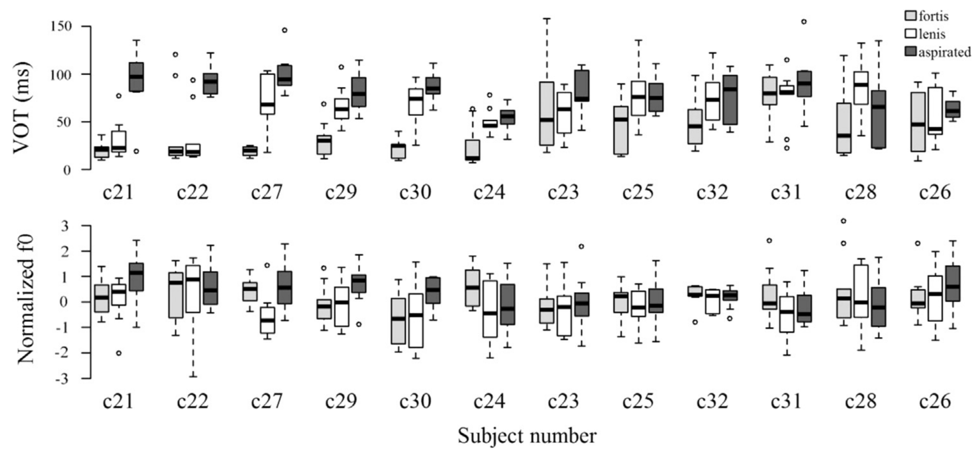

Fig. 2. Boxplots of VOT and *f*0 for each Korean velar phonation type for each subject. The subjects are in the same order as in Fig. 1.

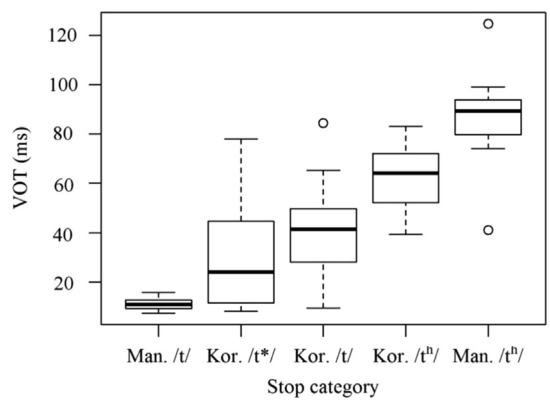

Fig. 3. By-subject mean VOT for each alveolar stop category in both Mandarin and Korean.

2.2.3. Comparison of L1 Mandarin and L2 Korean alveolar stops

Our next question was how well the VOT of the subjects' Korean stop categories mapped onto their L1 Mandarin stop categories. Previous studies of perceptual assimilation by naïve L1 Mandarin listeners suggested that Korean fortis stops are consistently assimilated to Mandarin unaspirated stops and Korean lenis and aspirated stops are consistently assimilated to Mandarin aspirated stops (Holliday, 2014). The current analysis aimed to investigate whether this pattern is realized in novice L2 production. This analysis was restricted to only the Korean alveolar stops because only alveolar stops were elicited in Mandarin and the difference in place of articulation would make a comparison with the Korean velar stops difficult to interpret (Lisker & Abramson, 1964).

Because VOT was the only dependent variable, a one-way ANOVA was run with a within-subjects factor of stop category. The factor had five levels, corresponding to the two Mandarin stop categories (unaspirated and aspirated) and three Korean stop categories. The effect of stop category was statistically significant ($F(4,44)=51.68$, $p<.001$), indicating that there were differences in VOT among the five stop categories. The means of the Mandarin stop categories were 11.1 ms for Mandarin /t/ and 86.6 ms for Mandarin /t^h/ (cf. 31.4 ms for Korean /t*/, 41.2 for Korean /t/, and 62.0 for Korean /t^h/). Tukey post-hoc comparisons among the five stop categories indicated that all pairwise comparisons were significant ($p<.001$, except for fortis-lenis $p=.010$). The distributions of by-subject by-category mean VOTs are shown in Fig. 3, which indeed suggest a five-way VOT contrast in line with the post-hoc results.

The individual VOT patterns shown in Fig. 4, however, paint a different picture. There are some subjects, such as c23, c25, and c32, whose data approximate the group trend of differentiating between all five categories. It seems that the entire first row of subjects, c21, c22, c27, c29, c30, and c24, all produced the Korean fortis stops as exemplars of the Mandarin unaspirated stop category, a pattern predicted by the perceptual assimilation patterns reported in Holliday (2014). Subjects c21 and c22 also produced the Korean lenis stops in the same way. The entire bottom row of subjects, c23, c25, c32, c31, c28, and c26, did not appear to produce Korean fortis stops as exemplars of Mandarin unaspirated stops, and in the case of subjects c31 and c28, produced all three Korean stop categories with VOT similar to that of the Mandarin aspirated stop category.

2.3. Discussion

The results of Experiment 1 revealed several trends in the Korean stop productions of L1 Mandarin learners of Korean. First, even though group-wide statistical analyses seemed to suggest that the subjects had implemented a three-way VOT contrast between

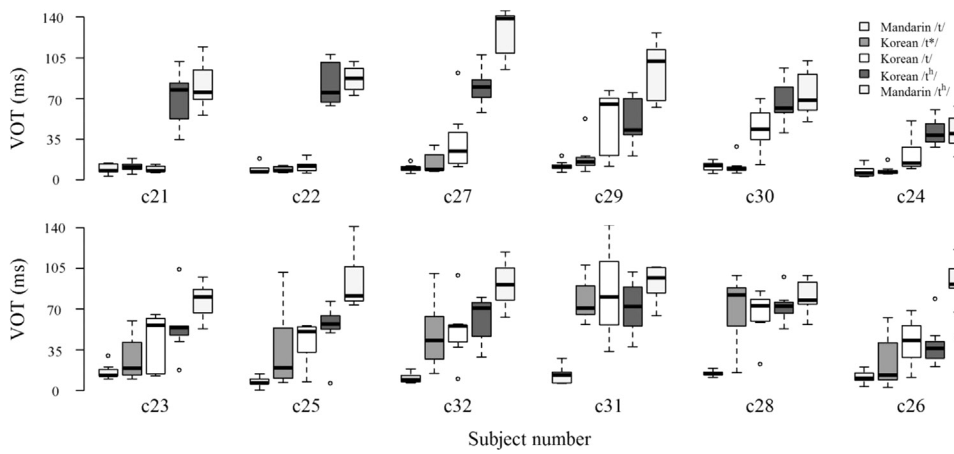

Fig. 4. Mean VOT for each Mandarin and Korean alveolar stop category for each subject. The order of bars from left to right is Mandarin unaspirated /t/, Korean fortis /t*/, Korean lenis /t/, Korean aspirated /t^h/, and Mandarin aspirated /t^h/. The subjects are in the same order as in Figs. 1 and 2.

Korean fortis, lenis, and aspirated stops, inspection of individual subjects' productions revealed a variety of different patterns. Some subjects' fortis and lenis stops clustered together, whereas other subjects' lenis and aspirated stops clustered together, other subjects seemed to differentiate between all three, and other subjects appeared to not differentiate between any of the phonation types.

This variety of individual patterns was initially unexpected in light of predictions made by the SLM (Flege, 1995) and the results of Holliday (2014). Because naïve L1 Mandarin listeners primarily use VOT in perceiving word-initial Korean stops (e.g. short-lag Korean fortis stops are perceived as short-lag Mandarin unaspirated stops), we should expect novice L2 learners of Korean to perceive them the same way, and to produce them with VOTs in line with the Mandarin stop categories they are perceived as. The variety of VOT patterns in the subjects' productions demonstrate that novice L2 stop production cannot be accurately predicted by naïve L1 perception alone and that perception may change as L2 acquisition proceeds, as claimed by PAM-L2 (Best & Tyler, 2007). These results are reminiscent of Chang (2010), who also reported substantial variation in how L1 English novice learners of Korean partitioned the acoustic space in their production of Korean stops, suggesting that the wide variety of production strategies found in Chang (2010) was not unique to L1 English learners.

Second, with the exception of subject c27, there was little evidence that subjects had learned to implement $f0$ correctly. The differences in $f0$ among phonation types for most subjects were much smaller than the differences typically produced by native Korean speakers (cf. Holliday & Kong, 2011; Kang & Guion, 2008), and almost none of the subjects exhibited the native-like pattern of producing lenis stops with $f0$ lower than that of fortis and aspirated stops. Although L1 Mandarin speakers should be sensitive to differences in $f0$, and have been shown to perceive phonetic differences between unfamiliar tones (e.g. Wu, Munro, & Wang, 2011), in the current study it could have been the case that the L1 Mandarin speakers were simply not aware that $f0$ was a relevant cue to the Korean stop contrast and therefore never paid attention to it.

Even if we ignore the L1 Mandarin speakers' failure to correctly implement $f0$ in their Korean stop productions, the wide variety of VOT patterns suggests that naïve perceptual assimilation patterns may have limited predictive power for L2 production. What remained to be seen was whether, with more L2 experience, these novice L2 learners would eventually converge on a single L1-influenced production strategy, or whether learners who seemed to cluster together at this novice stage would continue to pattern similarly. Related to the latter question, were there certain novice production patterns that might result in more native-like Korean production at a later stage than others? To answer these questions, a subset of the participants from Experiment 1 was recorded two more times, at approximately 6 and 12 months after the original recording.

3. Experiment 2

3.1. Methods

3.1.1. Participants

The participants were six L1 Mandarin learners of Korean (5 female, 1 male) who participated in Experiment 1. The remaining six participants from Experiment 1 were either unavailable to participate or did not respond when contacted.

3.1.2. Materials

The materials were identical to Experiment 1.

3.1.3. Procedure

The procedure was identical to Experiment 1 except that subjects were recorded twice, at approximately 6 and 12 months after their participation in Experiment 1. During the intervening period subjects had remained enrolled in the same intensive Korean language program, and had thus been receiving classroom Korean instruction for 4 h each day 5 days per week. For the remainder of the paper, the recordings from Experiment 1 will be referred to as “T1”, and the recordings done 6 and 12 months later will be referred to as “T2” and “T3”, respectively.

3.2. Results

3.2.1. Realization of Korean stop contrast at T2 and T3

The statistical analyses for the T2 and T3 recordings were done in the same way as for T1: four individual MANOVAs for the T2 and T3 alveolar and velar stops with dependent variables of VOT and $f0$ and independent variables of phonation type and subject. The results, shown in Table 3, indicate that phonation type was a statistically significant predictor for all four models. Individual ANOVAs were then run for both VOT and $f0$ for each of the four data sets (e.g. T2 and T3 alveolar and velar stops), which revealed that both VOT and $f0$ differed significantly across phonation types in all four models.

The interaction between phonation type and subject was statistically significant for each of the VOT models and the $f0$ models at T3, but not in the case of $f0$ at T2, indicating that there was generally significant variation between subjects. Given the large individual variation in the use of VOT and $f0$ at T1, the significant interactions between phonation type and subject at T2 and T3 suggest that the learners' production patterns did not converge even after either 6 or 12 months of L2 experience. Accordingly, the remaining analysis investigated how individual learners' productions changed over time, rather than try to generalize across learners at each time point.

3.2.2. Intra-speaker change over time

To investigate changes in VOT and $f0$ usage over time within each individual speaker, individual three-way ANOVAs were run for each speaker with phonation type, place of articulation, time point, and all possible interactions as independent variables and either VOT or $f0$ as the dependent variable. After ANOVAs were run for both VOT and $f0$ for each speaker, post-hoc Tukey HSD tests were then run on each model that showed any significant effects to determine which differences were statistically significant.

The VOT ANOVA results for each speaker are given in Table 4. All speakers had statistically significant effects of both phonation type and place of articulation on VOT, indicating that over all time points there were significant differences between at least one phonation type and another and between the alveolar and velar stops. The difference between alveolar and velar stops was expected, as this effect has been well-demonstrated across languages (Lisker & Abramson, 1964). Of most interest to the current study, subjects c22, c27, c29, and c32 had significant interactions between phonation type and time point, indicating that the effect of phonation type on VOT changed over time.

Fig. 5 shows how the distribution of VOT changed over time for each phonation type for each subject. Alveolar and velar stops have been combined in Fig. 5 because there were no statistically significant interactions between place of articulation and time point

Table 3
Wilks' lambda from MANOVAs of T2 and T3 data, and output from the ANOVAs with VOT and f0 as the dependent variable.

	MANOVA (Wilks' lambda)	ANOVA	Main effect	Subject interaction
T2 alveolar	$F(4,214)=32.17^{***}$	VOT	$F(2,108)=50.9^{***}$	$F(10,108)=3.2^{**}$
		$f0$	$F(2,108)=12.5^{***}$	$F(10,108)=1.8^{n.s.}$
T3 alveolar	$F(4,214)=49.40^{***}$	VOT	$F(2,108)=73.3^{***}$	$F(10,108)=2.6^{**}$
		$f0$	$F(2,108)=21.9^{***}$	$F(10,108)=3.7^{***}$
T2 velar	$F(4,322)=39.02^{***}$	VOT	$F(2,162)=90.2^{***}$	$F(10,162)=6.7^{***}$
		$f0$	$F(2,162)=3.40^{*}$	$F(10,162)=1.4^{n.s.}$
T3 velar	$F(4,322)=51.68^{***}$	VOT	$F(2,162)=95.4^{***}$	$F(10,162)=5.3^{***}$
		$f0$	$F(2,162)=19.1^{***}$	$F(10,162)=2.6^{**}$

* $p<.05$.
** $p<.01$.
*** $p<.001$.

Table 4
ANOVA output for VOT showing all factors with statistically significant effects.

	Subject					
	c21	c22	c27	c29	c25	c32
Phonation type	74.553^{***}	77.728^{***}	198.91^{***}	114.31^{***}	39.131^{***}	43.067^{***}
Place of articulation	15.877^{***}	16.325^{***}	30.984^{***}	18.754^{***}	14.976^{***}	23.403^{***}
Time point	3.579^{*}	3.429^{*}	12.681^{***}	13.254^{***}	2.26	3.374^{*}
Phonation × Place	.628	2.418	7.935^{***}	.586	5.643^{**}	4.152^{*}
Phonation × Time	.533	3.057^{*}	4.352^{**}	4.026^{**}	2.422	3.150^{*}

* $p<.05$.
** $p<.01$.
*** $p<.001$.

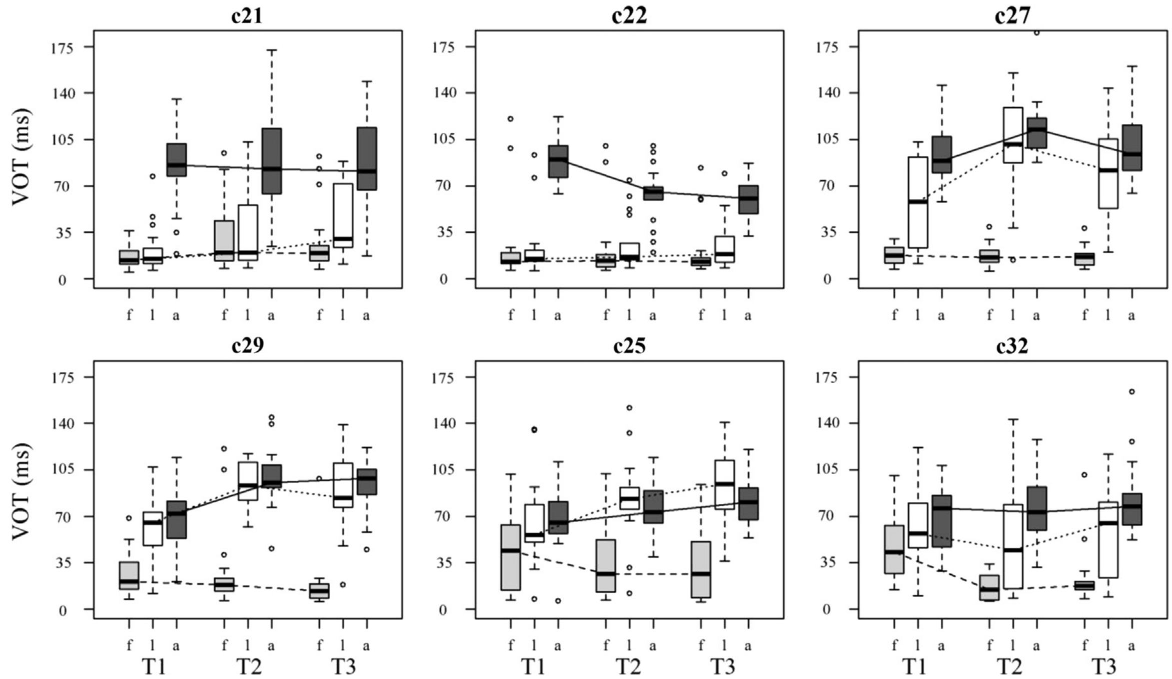

Fig. 5. The distribution of VOT for each phonation type at each time point for each subject. The dashed, dotted, and solid lines connect the medians of the fortis (f), lenis (l), and aspirated (a) stops, respectively, across all three time points.

or between phonation type, place of articulation, and time point. Although some subjects had an interaction between place of articulation and phonation type, because there was no subsequent interaction with time point it can be assumed that this interaction was consistent across all time points. Each time point is represented by three boxplots, one each for fortis (f), lenis (l), and aspirated (a) stops. The lines in each plot connect the medians from each box of the same phonation type, to help illustrate how the median value changed over time.

First, subjects c21 and c22 patterned very similarly to each other, with statistically significant effects of phonation type, place of articulation, and time point. Post-hoc Tukey HSD tests revealed that the statistically significant differences in VOT for either of these subjects was the comparison between T1 and T3 for subject c22 ($p = .038$). Thus the significant main effect of time point shown in Table 4 for these two subjects may not reflect significant system-wide shifts in VOT over time.

Tukey HSD tests for both subjects c21 and c22 also revealed that fortis and lenis stops did not differ significantly from each other, and that both phonation types differed significantly from aspirated stops. Although subject c22 had a significant interaction between phonation type and time point, it was found that the only significant comparisons were between aspirated stops at any time point and both lenis and fortis stops at any time point, in addition to the difference between aspirated stops at T1 and at both T2 and T3 (all comparisons $p < .001$). Put another way, subject c22's aspirated stops were significantly different from his lenis and fortis stops at all time points, and his lenis and fortis stops were never significantly different from each other. Thus, the significant interaction between phonation type and time point in Table 4 was due only to the VOT of subject c22's aspirated stops decreasing significantly between T1 and T2. For subject c21, there was no interaction whatsoever between phonation type and time point reflected in VOT. In summary, both subjects c21 and c22 can be described as having produced a non-native-like two-way VOT contrast between short lag fortis and lenis stops and long lag aspirated stops that remained unchanged after 12 months of L2 instruction.

Subject c27 showed statistically significant interactions between both phonation type and place of articulation and between phonation type and time point. Post-hoc Tukey HSD tests showed that subject c27 made an overall three-way VOT distinction between phonation types, but a closer look at the interaction between phonation type and time point revealed this was only true at T1. At T2 and T3, fortis stops still differed from both lenis and aspirated stops (both comparisons $p < .001$), but the difference between lenis and aspirated stops was not significant at either T2 ($p = .657$) or T3 ($p = .203$). The main effect of time point shown in Table 4 was the result of an overall increase in VOT from T1 to T2 and a subsequent decrease from T2 to T3. Thus, subject c27 started by distinguishing all three phonation types from each other, and later merged the VOT of lenis and aspirated stops to produce a two-way VOT distinction between short lag and long lag stops.

Subject c29 differed from the other five subjects in that at T1 she had already exhibited a native-like distinction between short lag fortis stops and long lag lenis and aspirated stops, confirmed by the post-hoc tests (lenis vs. aspirated $p = .939$; fortis vs. lenis $p = .002$; fortis vs. aspirated $p < .001$). The same two-way distinction was found at both T2 and T3, but with increased distance between the fortis stops and lenis and aspirated stops, as shown in Fig. 5.

Subject c25 showed statistically significant effects of phonation type and place of articulation and an interaction between these two factors, but no significant effect of time point or any interaction between time point and either phonation type or place of articulation. The interaction between phonation type and time point approached significance ($p = .051$), however, and visual inspection of the change in subject c25's stop productions over time in Fig. 5 does indeed suggest some separation between fortis stops and lenis and aspirated stops over time. Post-hoc Tukey HSD tests showed that the differences between the three phonation types were not significant at T1, but that fortis stops differed from lenis and aspirated stops at both T2 and T3 (all comparisons $p<.003$) and lenis and aspirated stops did not differ from each other at either T2 or T3 (both comparisons $p>.950$).

Lastly, subject c32 was similar to subject c25 in that at T1 she showed no distinction among the three phonation types whatsoever in terms of VOT (all comparisons $p>.250$). At T2, subject c32 had transitioned into the correct two-way VOT distinction between short lag fortis stops and long lag lenis and aspirated stops, although the difference between the lenis and aspirated stops at T2 approached significance ($p=.054$). By T3, the difference between lenis and aspirated stops was far from significant ($p=.350$), but the fortis stops were significantly different from both the lenis and aspirated stops (both comparisons $p\leq.001$). The trajectory of subjects c25 and c32 were thus quite similar, the main difference being that subject c25 seemed to have a more robust two-way VOT distinction at T2 than subject c32 did.

It should be noted that none of the subjects exhibited any meaningful shifts in the VOT of their Mandarin stop categories over the course of the year with the exception of subject c27. A two-way ANOVA with VOT as the dependent variable and Mandarin stop category and time point as independent variables yielded main effects of not only stop category but also time point [$F(2,48)=3.59$, $p=.035$]. Post-hoc Tukey HSD tests revealed that this effect was driven by a decrease in VOT of the Mandarin aspirated stop category, which differed from T1 to T3 ($p=.019$), but not between T1 and T2 or between T2 and T3. For all other subjects, the only statistically significant main effect was stop category, indicating that all subjects continued to produce a robust difference between Mandarin unaspirated and aspirated stops.

Turning next to the results for *f*0, the ANOVA output in Table 5 shows that overall subjects made a smaller distinction between phonation types than they did for VOT. First, only four subjects showed a main effect of phonation type, and of those only two showed a significant interaction between phonation type and time point. As expected, place of articulation accounted for very little variance, and no subjects showed an interaction between phonation type and place of articulation.

Table 5
ANOVA output for *f*0.

	Subject					
	c21	c22	c27	c29	c25	c32
Phonation type	8.795***	1.221	89.591***	18.726***	20.321***	.26
Place of articulation	.801	3.369	4.565*	2.8	.744	1.394
Time point	6.056**	4.622*	12.842***	10.690***	3.207*	1.327
Phonation × Place	.007	1.173	2.983	.746	.114	.086
Phonation × Time	.233	.212	.867	3.606**	2.515*	.349

* $p<.05$.
** $p<.01$.
*** $p<.001$.

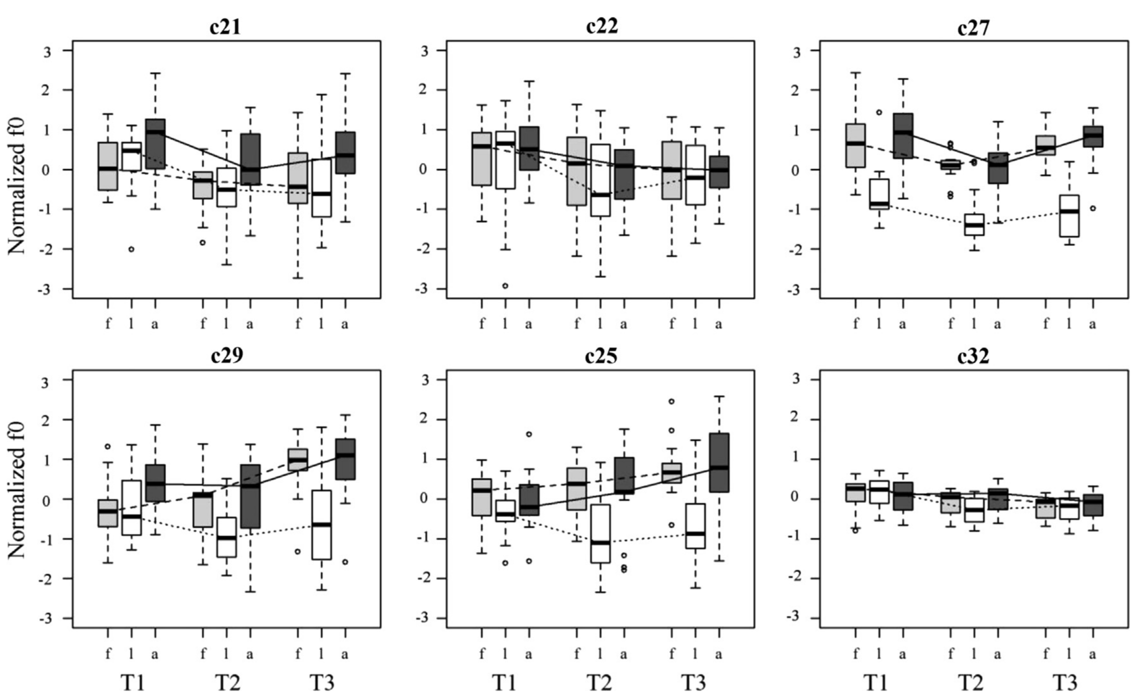

Fig. 6. The distribution of *f*0 for each phonation type at each time point for each speaker. The dashed, dotted, and solid lines connect the medians of the fortis (f), lenis (l), and aspirated (a) stops, respectively, across all three time points.

The distribution of $f0$ values across all speakers, phonation types, and time points is shown in Fig. 6. First, subjects c22 and 32 show very little differentiation between phonation types. Subject c22 had an overall statistically significant decrease in $f0$ from T1 to T2 ($p=.011$), but no differences between any phonation types at any time point, and subject c32 had no difference between any phonation type either overall or at any individual time point. These two subjects do not appear to have acquired the $f0$ contrast associated with Korean spots, either in full or in part.

Subjects c21 and c27 showed statistically significant main effects of phonation type and time point but no significant interaction between phonation type and time point. Subject c21, according to post-hoc Tukey HSD tests, showed an overall difference between aspirated stops and both fortis ($p=.001$) and lenis ($p=.001$) stops, with aspirated stops having a higher $f0$ than the latter two. This difference was not statistically significant at any individual time point, however, and fortis and lenis stops never differed from each other in terms of $f0$. Thus, for subject c21, the lack of a significant interaction between phonation type and time point indicates that she was unsuccessful in learning the $f0$ contrast, like subjects c22 and c32. Subject c27, on the other hand, produced the correct distinction between low $f0$ lenis stops and high $f0$ fortis and aspirated stops at all time points (all comparisons $p<.001$), suggesting that the lack of a significant interaction between phonation type and time reflected the fact that subject c27 produced the contrast correctly from the beginning.

The remaining two subjects, c25 and c29, both showed evidence of having acquired the $f0$ contrast by T3, if not earlier. Subject c25 did not differentiate between any phonation types at T1, but at T2 correctly differentiated between lenis and fortis stops ($p=.022$) and lenis and aspirated stops ($p=.009$), and continued to do so at T3 (both comparisons $p<.001$). Subject c29 patterned very similarly to subject c25, with the same overall distinction between low $f0$ lenis stops and high $f0$ fortis and aspirated stops. Like subject c25, no distinction between phonation types was made at T1, and at T2 subject c29 only produced a significant difference in $f0$ between lenis and aspirated stops ($p=.027$) but not between lenis and fortis stops. At T3, however, subject c29 produced the stop contrast with native-like $f0$ correlates, with significant differences between lenis stops and both fortis and aspirated stops (both comparisons $p<.001$).

4. General discussion

4.1. Apparent mismatch between perception and production

This study presented evidence that L2 learners acquiring a non-native phonological contrast that relies on two primary acoustic cues can vary widely in how the acoustic cues are implemented, even when one of the cues (in this case, VOT) is the primary cue for a similar L1 contrast. This finding was similar to Chang (2010), who reported a comparable level of variation across L1 English learners of Korean, and also noted that the variety of production patterns did not match up with the perception results reported in Schmidt (2007). The same can be said of the results of the current study. If we follow the claim that L2 production targets are guided by perception (Flege, 1995), the results of Holliday (2014) predict that L1 Mandarin learners of Korean should produce Korean fortis stops with short lag VOT and Korean lenis and aspirated stops with long lag VOT, but only as long as the L2 learners' perception of Korean stops does not differ from that of naïve listeners.

PAM-L2 (Best & Tyler, 2007) resolves this apparent inconsistency. If we accept that perceptual assimilation patterns can change as L2 learning proceeds, then we are not obliged to reconcile the production patterns of L2 learners with the perception patterns of naïve listeners. This line of reasoning implicates the need for more studies comparing perceptual assimilation between naïve listeners and L2 learners, but unfortunately very few studies have done so. One such study, Levy (2009), showed that L1 English naïve listeners and L2 learners differed in their perceptual assimilation of certain French vowels. Wagner and Baker-Smemoe (2013), on the other hand, reported no differences between L1 English naïve listeners and L2 learners in their perceptual assimilation of Q'eqchi' Mayan stops, despite the fact that the L2 learners in that study were very experienced and had lived in an immersion environment. Other studies have tested perceptual assimilation in L2 learners (e.g. Bundgaard-Nielsen, Best, & Tyler, 2011; Hattori & Iverson, 2009; Mayr & Escudero, 2010), but did not provide a direct comparison with naïve listeners. PAM-L2 does not claim that the perceptual assimilation patterns of naïve listeners and L2 listeners must necessarily differ, but they should to the extent that we find L2 production patterns that are not predicted by naïve perception. Thus, a study comparing the perceptual assimilation of Korean stops by both naïve listeners and L2 learners with the same experimental paradigm could help in the interpretation of the results of the current study.

Among the explanations offered by Chang (2010) for the wide variety of production patterns found in his results, the influence of orthography is one that could possibly explain rapid changes in perceptual assimilation. Korean is written in a non-Latin orthography, called Hangul, but Korean language textbooks for L2 learners usually introduce Hangul alongside a quasi-IPA Latin transliteration meant to aid the learner. For example, in the textbook used by the participants in the current study, fortis stops were represented as "t_*", lenis stops as "t/d", and aspirated stops as "t^h" (Korean Language & Culture Center, 2008). In addition, most public signage in Korea (e.g. street signs, public transportation signs, names of landmarks, etc.) is rendered in both Korean and English, but place names are usually transliterated instead of translated. This situation results in very novice L2 learners of Korean being exposed to and, crucially, having to rely on transliterated Korean, as at that stage they are still learning to read Hangul. Although Mandarin is written in Chinese characters, speakers from mainland China are also taught a Romanized form called pinyin in which Mandarin short lag unaspirated stops correspond to <b, d, g> and Mandarin long lag aspirated stops correspond to <p, t, k>.

In the transliteration system used by the Korean government (National Institute of the Korean Language, 2014), word-initial aspirated stops are written as <p, t, k>, lenis stops as <b, d, g>, and fortis stops as <pp, tt, kk>. This system could prove confusing to an L1 Mandarin or L1 English learner of Korean, who would naïvely (that is, auditorily) assimilate Korean fortis stops to <b, d, g> and lenis stops to <p, t, k> (for L1 English, Schmidt, 2007; for L1 Mandarin, Holliday, 2014). Orthography has been shown to influence L1 perception and production in subtle ways (e.g. Fourakis & Iverson, 1984; Tanenhaus, Flanigan, & Seidenberg, 1980; Warner, Jongman, Sereno, & Kemps, 2004), and recent studies have shown that exposure to orthography can bias the perception and production of L2 learners, as well (Escudero & Wanrooij, 2010; Escudero, Hayes-Harb, & Mitterer, 2008; Ota, Hartsuiker, & Haywood, 2010). This sort of interference could help explain the pattern, found most strongly in subjects c21 and c22, of producing Korean lenis stops with either very short lag VOT or at least shorter lag VOT than aspirated stops. These subjects still produced Korean fortis stops with short lag VOT, which could contradict an orthographic explanation if indeed these subjects thought that orthographic <pp, tt, kk> corresponded to pinyin <p, t, k>. It is equally plausible, however, that because <pp, tt, kk> do not occur in pinyin their novelty could have discouraged learners from relying on their orthographic forms in the same way that <b, d, g> might encourage learners to map Korean lenis stops onto Mandarin unaspirated stops.

Another reason why lenis stops might be susceptible to orthographic influence is that their phonetic realization is more varied than the other two phonation types: in word-initial position they are realized with long lag VOT, in intervocalic position they are voiced, and in post-obstruent position they are realized as fortis stops. Korean fortis and aspirated stops, however, are produced with short and long lag VOT, respectively, in all non-coda environments. Novice L2 learners of Korean, therefore, could be confused by the variety of ways in which lenis stops are phonetically realized, which might allow their orthographic representation to bias their perception. Further enforcing the correspondence between Korean lenis stops and Mandarin unaspirated stops is the fact that two of these phonetic realizations, intervocalically voiced lenis stops and post-obstruent fortified lenis stops, would be perceived by L1 Mandarin listeners as unaspirated anyway. Taken together, all of these possibilities suggest that, in line with the SLM (Flege, 1995), Korean lenis stops could be produced with short lag VOT because their perceptual targets are closer to L1 Mandarin speakers' short lag VOT unaspirated stops.

4.2. Effect of initial strategy on long-term success

One of the most striking results from the longitudinal analysis presented in Experiment 2 was that two subjects, c21 and c22, produced Korean lenis stops with short lag VOT very consistently at all three time points, showing no evidence of having learned the correct VOT contrast. In the previous section some reasons were offered why an L2 learner might produce Korean lenis stops with short lag VOT to begin with, but it is not clear why only subjects c21 and c22 persisted in their initial production strategy. An additional quantitative difference between these two subjects and the other four in the longitudinal study is that subjects c21 and c22 exhibited much less within-category variation in the VOT of their lenis stop productions. The boxplots in Figs. 4 and 5 make it clear that subjects c21 and c22 were not unsure of how to produce lenis stops: they were produced with a VOT distribution almost identical and overlapping with both fortis stops and their L1 Mandarin unaspirated stops. The other subjects' lenis stop productions had more variable VOT, with standard deviations for their alveolar productions ranging from 18 to 29 ms. Subjects c21 and c22, on the other hand, had standard deviations of 3 and 5 ms, respectively.

There has been very little work on the effect of within-speaker variability on the development of L2 speech production, partly because there have not been many longitudinal L2 production studies in the first place. The patterns observed in the current study, however, have parallels in the L1 acquisition literature. Early studies of covert contrast in children drew a link between covert "productive knowledge" of the target (in this case, L1) phonology, manifested through measurable but imperceptible acoustic differences between members of a contrast, and intervention treatment outcomes (e.g. Dinnsen & Elbert, 1984; Gierut & Dinnsen, 1986; Tyler, Edwards, & Saxman, 1990; Tyler et al., 1993). It was generally found that children who showed evidence of "productive knowledge" in the form of a covert contrast had more positive treatment outcomes than children who showed no evidence of contrast whatsoever. A possible analogy to the current study is that the L2 learners who exhibited the least amount of flexibility in their lenis stop productions at T1, namely c21 and c22, were the least successful learners one year later. The two other subjects who also exhibited incorrect production at T1, c25 and c32, were far less consistent, suggesting that they had not adopted any strategy in particular, and eventually learned the correct VOT contrast. Thus, both these previous studies and the current study jointly address the idea that flexibility in a learner's productions may predict later success in both L1 and L2 speech development.

4.3. Acquisition of the f0 cue

It was initially predicted that L1 speakers of Mandarin, a tonal language, would acquire the *f*0 cue to the Korean stop contrast somewhat easily, and perform at least as well as the L1 English learners in Chang (2010). At T1, no group-wide statistically significant difference in *f*0 was found among any of the alveolar stops, and for the velar stops a group-wide significant *f*0 difference was found only between lenis /k/ and aspirated /k^h/. Inspection of individual trends revealed one subject, c27, who produced a native-like *f*0 contrast from T1, but the other five subjects in the longitudinal study produced no *f*0 contrast whatsoever at T1. By T3, two other subjects, c25 and c29, had acquired the *f*0 contrast, but the other three subjects, c21, c22, and c32, had not learned to produce a reliable *f*0 difference among any of the phonation types after an entire year of in-country language instruction.

The results of the current study suggest that being a native speaker of a tone language may not confer any benefit in learning an *f*0-cued consonant contrast. The *f*0 difference between Korean lenis and aspirated stops has been shown to be quite large (e.g.

68 Hz in conversational speech, Kang & Guion, 2008) and so it is improbable that L1 speakers of Mandarin, or any language, for that matter, would be unable to perceive such differences under the right conditions. Thus, the data in this study seem to address the question of whether L1 Mandarin speakers *do* learn the Korean *f*0 stop cue in a typical L2 Korean classroom, and not whether they might learn it quite easily in a controlled laboratory environment.

5. Conclusion

The current study addressed the question of how L1 speakers of Mandarin, a language with a two-way VOT stop contrast, acquire the Korean three-way stop contrast, which is jointly cued by VOT and *f*0. While significant group-wide patterns were found in their use of VOT to cue the three-way contrast, inspection of individual production patterns revealed a wide variety of production strategies, some of which could not have been predicted by L1–L2 acoustic correspondences. It was also found that most of the novice L2 learners did not pick up on the *f*0 stop cue, contrary to the expectation that native speakers of a tone language would be especially sensitive to the *f*0 differences among the three phonation types. In a longitudinal analysis of a subset of the learners, it was found that the learners whose Korean stop productions were incorrect but consistent fared the worst after one year of L2 learning, and the learners whose productions were incorrect but variable demonstrated significant improvement after one year. Finally, it was found that only three of the six learners who participated in the longitudinal study demonstrated acquisition of the *f*0 stop cue after one year of L2 learning.

Part of the difficulty in acquiring the Korean three-way stop contrast could stem from the fact that the contrast relies simultaneously on two different acoustic dimensions. The learners in the present study seemed to be more successful in acquiring the VOT cue than the *f*0 cue, which raises the question of whether a three-way stop contrast that is cued primarily by VOT alone, such as Thai, would be easier for L1 Mandarin learners to acquire. Only having to manipulate a single acoustic cue could make it easier, but the VOT dimension itself must be split three ways instead of just two, as it is in modern Korean, which could also prove challenging. Investigating the L2 acquisition of both Korean and Thai stops by listeners from the same L1 population would shed light on how L2 learners manage different types of cue learning: splitting a familiar acoustic dimension into three categories instead of two, or using a familiar acoustic cue in a familiar way but having to manipulate a second cue simultaneously. It has been shown that naïve L1 English speakers can perceptually learn such systems with laboratory training (e.g. Francis & Nusbaum, 2002), but there have been very few production studies of actual L2 learners.

Lastly, the apparent mismatch between perception and production documented both here and in Chang (2010) raises an additional set of questions that cannot be answered by production data alone. Can the perceptual assimilation of certain L2 sounds change so dramatically after only a few weeks of L2 learning? It has been suggested that perceptual assimilation may not change with L2 learning (Wagner & Baker-Smemoe, 2013). Perceptual assimilation patterns may instead depend not only on factors such as the L1–L2 acoustic correspondences and the individual segments under investigation, but also on other factors such as orthography. In a recent study of L1 Korean speakers' English production, Hong, Kim, and Chung (2014) found that the vast majority of certain errors in vowel production were due to orthographic influence (e.g. producing the first vowel in <project> as [oʊ]). If such errors are spelling pronunciations, and not reflective of phonetic perception, then models of L2 speech production would need to account for such extra-auditory sources of bias.

Acknowledgments

The author would like to thank the Korean Language and Culture Center at Korea University for help in securing space and recruiting participants, and Stephanie Dickinson, Huizi Xu, and Zhili Xu from the Indiana Statistical Consulting Center for help with the statistical analyses. The author would also like to thank the following people for helpful comments over the course of this project: Mary Beckman, Kelly Berkson, Cynthia Clopper, Isabelle Darcy, Ken de Jong, Rachel Hayes-Harb, Elizabeth McCullough, and the audiences of the 32nd Second Language Research Forum and the Interspeech 2013 satellite workshop "Phonetics, Phonology and Language Contact." This research was supported by a Targeted Investment for Excellence Fellowship from the Ohio State University Department of Linguistics, an International Affairs Grant from the Ohio State University Office of International Affairs, and a Doctoral Dissertation Research Improvement Grant from the National Science Foundation (BCS-1024286) awarded to Mary Beckman and the author.

References

Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P.C. M. (2011). Music-melody perception in tone-language and non-tone-language speakers. In *Proceedings of the psycholinguistic representation of tone conference*.

Altenberg, E. P., & Ferrand, C. T. (2006). Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice*, *20*(1), 89–96.

Aoyama, K., Guion, S. G., Flege, J. E., Yamada, T., & Akahane-Yamada, R. (2008). The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English. *International Review of Applied Linguistics*, *46*, 61–90.

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

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

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer. Version 5.3.40. ⟨http://www.praat.org⟩.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(1), 51–67.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *Journal of the Acoustical Society of America*, *54*(2), 421–428.

Chang, C. B. (2010). The implementation of laryngeal contrast in Korean as a second language. In Susumu Kuno, Ik-Hwan Lee, John Whitman, Joan Maling, Young-Se Kang, Peter Sells, Hyang-Sook Sohn, & Youngjun Jang (Eds.), *Harvard studies in Korean Linguistics XIII* (pp. 91–104). Seoul, South Korea: Hanshin.

Dinnsen, D. A., & Elbert, M. (1984). On the relationship between phonology and learning. In M. Elbert, D. A. Dinnsen, & G. Weismer (Eds.), *Phonological theory and the misarticulating child* (pp. 59–68). Rockville, MD: ASHA.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, *53*(3), 343–365.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89* (1), 395–411.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233–277.

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

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, *83*(2), 729–740.

Forrest, K., Dinnsen, D. A., & Elbert, M. (1997). Impact of substitution patterns on phonological learning by misarticulating children. *Clinical Linguistics and Phonetics*, *11*(1), 63–76.

Fourakis, M., & Iverson, G. K. (1984). On the 'incomplete neutralization'of German final obstruents. *Phonetica*, *41*(3), 140–149.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics*, *36*(4), 649–663.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28* (2), 349–366.

Gierut, J. A., & Dinnsen, D. A. (1986). On word-initial voicing: Converging sources of evidence in phonologically disordered speech. *Language and Speech*, *29*(2), 97–114.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

Han, M. S., & Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/. *Phonetica*, *22*, 112–128.

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

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French–English bilinguals. *Language and Speech*, *36*(1), 17–38.

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

Holliday, J.J., & Kong, E.J. (2011). Dialectal variation in the acoustic correlates of Korean stops. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 878–881).

Hong, H., Kim, S., & Chung, M. (2014). A corpus-based analysis of English segments produced by Korean learners. *Journal of Phonetics*, *46*, 52–67.

Jung, M., & Kwon, S. (2010). A study on the phonetic discrimination and acquisition ability of Korean language learners. *Journal of the Korean Society of Speech Sciences*, *2*(1), 23–32.

Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics*, *2*, 161–180.

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

Kang, K.-H., & Guion, S. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, *124*(2), 3909–3917.

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.

Korean Language & Culture Center. (2008). *Caymi issnun hankwuke [Fun! Fun! Korean]*. Seoul: Kyobo.

Ladefoged, P. (1999). *American English*. *Handbook of the International Phonetic Association: A guide to the use of the international phonetic alphabet* (pp. 41–44)Cambridge: Cambridge University Press41–44.

Levy, E. S. (2009). On the assimilation–discrimination relationship in American English adults' French vowel learning. *Journal of the Acoustical Society of America*, *126*(5), 2670–2682.

Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, *37*(1), 111–124.

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

Macken, M. A., & Barton, D. (1980). A longitudinal study of the acquisition of the voicing contrast in American-English word-initial stops, as measured by voice onset time. *Journal of Child Language*, *7*, 41–74.

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

National Institute of the Korean Language. (2014). Rules for Romanization. Retrieved from 〈http://www.korean.go.kr/09_new/dic/rule/rule_roman_0101.jsp〉.

Ng, M. L., Chen, Y., & Chan, E. Y. K. (2012). Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers – A long-term average spectral analysis. *Journal of Voice*, *26*(4), e171–e176.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, *39*(2), 156–167.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2010). Is a FAN always FUN? Phonological and orthographic effects in bilingual visual word recognition. *Language and Speech*, *53*(3), 383–403.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436.

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

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

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293.

Stölten, K., Abrahamsson, N., Hyltenstam, K. Effects of age and speaking rate on voice onset time: The production of voiceless stops by near-native L2 speakers. Studies in Second Language Acquisition, http://dx.doi.org/10.1017/S0272263114000151, in press.

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

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition*, *8*(6), 513–520.

Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., & Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica*, *61*(2–3), 67–83.

Tyler, A. A., Edwards, M. L., & Saxman, J. H. (1990). Acoustic validation of phonological knowledge and its relationship to treatment. *Journal of Speech and Hearing Disorders*, *55*(2), 251–261.

Tyler, A. A., Figurski, G. R., & Langsdale, T. (1993). Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech, Language, and Hearing Research*, *36*(4), 746–759.

Wagner, K. O. C., & Baker-Smemoe, W. (2013). An investigation of the production of ejectives by native (L1) and second (L2) language speakers of Q'eqchi'Mayan. *Journal of Phonetics*, *41*(6), 453–467.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*(2), 251–276.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712.

Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology, Volume 2: Perception*. New York: Academic Press.

Wu, X., Munro, M. J., & Wang, Y. (2011). The effects of linguistic experience on perceptual assimilation of lexical tone. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2169–2172).

Yasuta, T. (2004). *Stop perception in second language phonology: Perception of English and Korean stops by Japanese speakers* (Ph.D. thesis). University of Hawai'i, Manoa.