Tone-induced split in stop category mapping by Korean learners of Mandarin Chinese

Sang-Im Lee-Kim
National Chiao Tung University

The present study reports a novel case where a simple one-to-one category mapping may develop into a systematic one-to-two mapping over the course of second language acquisition. We examined the split in category mapping of the Mandarin unaspirated stops conditioned by tone by Korean-speaking learners of Mandarin Chinese (e.g. Mandarin /ta^{35}/ to Korean lenis [ta] vs. Mandarin /ta^{55}/ to Korean fortis [t'a]). Korean L2 learners and naïve listeners participated in identification tasks in which *f0* contours of Mandarin words containing unaspirated stops with short-lag VOTs were digitally manipulated. In word-initial position, learners showed a near-categorical perception from lenis to fortis as *f0* increased, while most stimuli were identified as fortis by naïve listeners. The effect of *f0* was much smaller in word-medial position, but the group difference remained the same, confirming the two groups' differential use of phonetic cues for stop identification. Taken together, a substantial reorganization of perceptual cues, namely the promotion of *f0* concurrent with significant underweighting of VOT cues, seems to have taken place during L2 acquisition. The findings were discussed with reference to PAM-L2 whereby the knowledge of the L2 phonological system along with particular phonetic properties of the L2 sounds may have driven a perceptual regrouping of the L2 stop categories.

Keywords: VOT, *f0*, stop category mapping, PAM-L2, perceptual reorganization

1. Introduction

The vast majority of the second language literature has demonstrated that learners of a second language strive to develop fine phonetic discriminability for difficult non-native contrasts through training (e.g. Bradlow et al. 1997; Francis et al. 2000) or practice (e.g. Flege 1987; Aoyama et al. 2004; Levy & Strange 2008) throughout

https://doi.org/10.1075/lali.00057.san
Language and Linguistics 21:1 (2020), pp. 54–79. issn 1606-822x | e-issn 2309-5067 © ILAS

one's lifespan. A prominent example comes from the English /ɹ-l/ discrimination by Japanese listeners. Many Japanese listeners fail to discriminate the English liquid contrast due to their over-reliance on an unreliable cue, F2, (Iverson et al. 2003), but this can be overcome by cue-specific training where the reliable F3 cue, for example, is specifically enhanced in the experimental stimuli (Iverson et al. 2005). Even without well-designed, laboratory-based training, however, Japanese listeners with sufficient experience become successful at figuring out the relevant cue through unsupervised everyday practice (Aoyama et al. 2004).

Yet more experience may not necessarily entail better identification or discrimination of foreign sounds and contrasts. As in the first language (L1) developmental path, some phonetic details may be disregarded over the course of the second language (L2) acquisition. A classic example can be found in the remarkable commonality in the ways in which L2 rhotic sounds are interpreted by the learners (Lindau 1985; Ladefoged & Maddieson 1996). Despite large phonetic differences or the near absence of acoustic resemblance between rhotics, learners tend to equate a foreign rhotic sound (e.g. French [ʁ~χ]) to their native rhotic sound (e.g. English /ɹ/). While the mapping does not make much sense from a phonetic perspective, learners are likely to tap into the abstract knowledge of the linguistic structures, namely the distributional similarity of rhotics in the two phonological systems. Some recent studies have shown that the departure from the acoustic-phonetic level processing by experienced learners may result in even poorer performance for some nonnative contrasts than naïve listeners. English L2 experienced learners of French were worse than inexperienced listeners at discriminating the French vowel contrast /u-y/, particularly in the bilabial context (Levy & Strange 2008). Similarly, Mandarin L2 learners of Korean performed more poorly at discriminating the Korean /s^{h}a-s*a/ contrasts than naïve listeners (Holliday 2016).

PAM-L2 (Best & Tyler 2007), a theoretical extension of the original Perceptual Assimilation Model (PAM, Best 1995), has been proposed to explain the potential changes in the perceptual assimilation of L2 sounds and contrasts by learners. Key to the PAM-L2 is the proposal that the knowledge of L2 forms and structures, not available to naïve listeners, may come into play to reshape how L2 sounds are interpreted. Sometimes the knowledge of L2 phonology can be so strong that heterogenetic phonetic details of sounds can be conveniently ignored, as exemplified in the rhotic mapping. The present study aims to contribute further empirical evidence to this growing body of literature by demonstrating that L2 experience may lead to a considerable divergence from the ways in which nonnative stops are interpreted by naïve listeners. The novel aspect of the current study is that the L2 phone in question is initially consistently assimilated to a single L1 phone, while this one-to-one mapping falls apart as one gains experience in

the L2. The particular case of interest is the perception of Mandarin unaspirated stops by Korean listeners with or without prior experience in Mandarin Chinese. Unlike most previous studies focusing on challenging cases where two non-native phones are assimilated to a single native phone (i.e. Single-Category Assimilation in PAM), we shall show that the straightforward one-to-one mapping may systematically develop into a one-to-two mapping through L2 acquisition.

1.1 Background

Korean has a typologically rare set of stop sounds in its inventory including fortis (/p' t' k'/), lenis (/p t k/), and aspirated (/p^h t^h k^h/). The three stops are categorized by a combination of phonetic cues, namely VOT and *f0* at the vowel onset among others, particularly in initial position (Lisker & Abramson 1964; Han & Weitzman 1970; Cho & Jun & Ladefoged 2002). Specifically, fortis stops manifest short VOTs and high *f0*. Lenis stops take the opposite values of both parameters, namely intermediate-to-long VOTs and low *f0*. Aspirated stops are characterized by long VOT and high *f0*. The rather clear-cut separation between the three stops along the VOT parameter is no longer observed in contemporary Seoul Korean mainly due to a remarkable VOT reduction in the aspirated stops (e.g. Silva 1992, 2006; Kang 2014). Meanwhile, *f0* differences at the vowel onset have increased significantly especially for the aspirated-lenis pair, compensating for the loss of the original VOT distinction.

While the increased role of *f0* in stop identification in general had certainly been acknowledged in perception studies (Kim et al. 2002; Kim 2004; Lee et al. 2013), the relative contribution of individual phonetic cues is subject to specific pairs under comparison. For the lenis-fortis contrast, in particular, short VOTs (less than 20 ms) are sufficient to exclusively signal fortis stops (Kim 2004; Lee et al. 2013). Lee et al. (2013), for instance, tested the trading relation between VOT and *f0* cues by systematically manipulating both parameters. Stop VOTs were manipulated to range between 10 and 142 ms, and *f0* to range between 99 and 209 Hz, both on a 12-step continuum. For stimuli with short VOTs, Seoul Koreans perceived the stops as fortis nearly exclusively *regardless* of *f0* (more than 90% for stimuli with VOTs of 10 ms and approximately 90% for stimuli with VOTs of 22 ms). When VOTs were manipulated to be intermediate (34 and 46 ms), fortis responses began to decrease and were limited to stimuli with higher *f0*. With VOTs longer than 58 ms, no fortis responses were made again regardless of *f0*.

An important implication of this finding is that participants' fortis judgment was not influenced by *f0* when the VOT cues were unambiguous. Seoul Koreans gave fortis responses predominantly when VOTs were sufficiently short, and the

fortis responses virtually disappeared for stimuli with sufficiently long VOTs. *F0* might come into play to influence stop perception only when the VOT cues were ambiguous ranging at intermediate values. Considering the acoustic characteristics of fortis stops, namely high *f0* and short VOTs, the results of the perception studies indicate that the perceptual weighting for fortis identification is asymmetrically and heavily placed on one of the major acoustic parameters, namely VOT.

Turning to the learners' second language, Mandarin Chinese, the language exhibits a common two-way contrast in which the stops are differentiated by aspiration: /p t k/ characterized by short VOTs vs. /p^h t^h k^h/ by long VOTs. Much like the one-to-one mapping of the English aspirated stops onto Korean aspirates (Schmidt 1996; Park & de Jong 2008), aspirated stops between Mandarin and Korean hold a straightforward correspondence both acoustically and perceptually. Mandarin aspirated stops have long VOT values (Rochet & Fei 1991; Liu et al. 2000; Chao & Chen 2008), and Mandarin-speaking learners of Korean are known to equate Korean aspirated stops to Mandarin aspirated ones perceptually (Kim 2006; Holliday 2014).

Similar to Korean fortis stops, unaspirated stops in Mandarin Chinese are characterized by short VOTs less than 20 ms, particularly for non-velar stops (Rochet & Fei 1991; Liu et al. 2000). Given the results of L1 perception studies demonstrating the dominance of VOTs for fortis perception, naïve Korean listeners are likely to interpret Mandarin unaspirated stops as fortis in Korean. This is, in fact, implemented as such in Chinese-Korean transcription conventions established by both Kim (1987) and Eom (1996) whereby Mandarin unaspirated stops are transcribed as fortis stops for Chinese words borrowed into Korean (e.g. [pa] in Mandarin transcribed as "빠 [p'a]" in Korean). To summarize, a quite straightforward one-to-one mapping holds between Mandarin and Korean stops: Mandarin aspirated stops onto Korean aspirates, and Mandarin unaspirated stops onto Korean fortis stops. According to PAM, this is a typical example of Two-Category Assimilation by which two foreign phones are assimilated to two L1 phones separately and, therefore, the discrimination of the aspirated and unaspirated stops in Mandarin is predicted to be easy for naïve Korean listeners.

What is puzzling is the variable mapping of the unaspirated stops conditioned by tone by *Korean L2 learners of Mandarin Chinese*. Previous studies have found that Mandarin unaspirated stops are mapped onto Korean fortis stops when the syllable carries high-level (X^{55}) or high-falling (X^{51}) tones, whereas they are mapped onto Korean lenis stops when the syllable carries mid-rising (X^{35}) or low-

dipping (X^{214}) tones (Ko 2000; Maeng & Kwon 2008), as schematized in Figure 1.[1] Maeng & Kwon (2008), for example, tested learners of Mandarin Chinese in perception experiments where *f0* of the syllables containing Mandarin unaspirated stops was manipulated to switch between high-level and mid-rising tones. The results showed that the learners' identification shifted to a great degree from lenis to fortis as the tone contour changed from mid-rising to high-level, and vice versa.

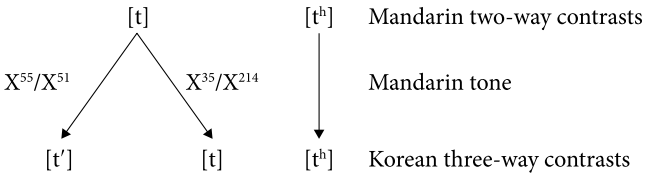

Figure 1. Perceptual mapping between Mandarin and Korean stops by Korean L2 learners of Mandarin Chinese

The particular grouping of Mandarin tones seems straightforward and the split seems systematic; in terms of *f0*, high-level and high-falling tone syllables start off with high *f0* (i.e. onset *f0* marked by 5), and mid-rising and low-dipping tone syllables begin with low *f0* (i.e. onset *f0* marked by 2 or 3). High onset *f0* appears to elicit fortis judgments, while relatively low onset *f0* appears to elicit lenis judgments. Nonetheless, the split in the perception of the Mandarin unaspirated stops by the learners does not make much sense given the particular cue-pairing found in native Korean. While the mapping of the Mandarin unaspirated stops in high-initial *f0* syllables (high *f0* – short VOT) onto Korean fortis stops (high *f0* – short VOT) creates a good correspondence along both phonetic parameters, Mandarin unaspirated stops in low-initial *f0* syllables (low *f0* – *short VOT*) being interpreted as Korean lenis stops (low *f0* – *long VOT*) brings a sharp discrepancy particularly for the VOT parameter.

While one might assume that the split of the unaspirated stops in the L2-to-L1 mapping is driven by a higher sensitivity to tonal properties of the L2 by the learners, it is not entirely clear whether the split can be solely attributed to the acquisition of the tonal second language. Naïve and experienced listeners have never been brought to the same experimental setting, and each group of listeners has only been tested on their corresponding languages; naïve listeners were tested on manipulated stimuli based on Korean (e.g. Kim 2004; Lee et al. 2013)

1. The tone letter system by Chao (1968) was adopted for the transcription of tones. In this convention, Mandarin four tones are transcribed with numerical pitch values ranging from the lowest 1 to the highest 5: [ma^{55}] for high-level, [ma^{35}] for mid-rising, [ma^{214}] for low-dipping, and [ma^{51}] for high-falling tone.

while Mandarin learners on stimuli based on Mandarin (Maeng & Kwon 2008). In particular, it is only a hypothesis that naïve Korean listeners are likely to interpret Mandarin unaspirated stops as Korean fortis, inferred from the experimental results from past research.

In order to test the effect of learning on the alleged dichotomy in stop category mapping among Korean listeners, it is warranted that the learners' behavior should be explicitly tested along with naïve listeners, who have not had any experience with tonal languages. If learners alone show the dominance of *f0* over other phonetic cues, it would provide evidence for the effect of learning on stop perception. In sum, the present study aims to establish a fuller picture of the perceptual patterns of the Mandarin unaspirated stops with varying *f0* by both groups of listeners in a single experiment. The learners' differential behavior, if any, will be discussed with reference to the PAM-L2, highlighting the role of the knowledge of L2 phonology on the split of L2 stop sounds.

1.2 The present study

The present study tested whether the learning of a tonal second language may lead to differential uses of phonetic cues in stop identification, departing from naïve Korean listeners. In doing so, three experimental variables were introduced in the identification tasks, including *f0*, word position, and tone contours. In order to test the differential effects of *f0*, target syllables with unaspirated stops selected from naturally produced Mandarin utterances were manipulated to vary in *f0*. The prediction is straightforward. The stop perception by the learners of Mandarin Chinese is expected to change as a function of *f0*, while naïve Korean listeners should be comparatively irresponsive to varying *f0*. The latter prediction, in particular, is based on the previous studies which showed that short-lag VOTs elicit fortis judgments nearly exclusively for Korean listeners. Therefore, special attention will be given to the interaction between learning experience and sensitivity to *f0* in the statistical analyses later on.

In addition to the identification of word-initial stops, word-medial stops were also included in the experimental design. The variation in word position was expected to help to draw a fuller picture of the effect of learning on stop perception. In particular, we capitalize on the remarkable change in the distribution of the acoustic cues to Korean stops in different word positions. In production, lenis stops are often subject to full or partial voicing in intervocalic position (e.g. Han & Weitzman 1970; Silva 1992; Han 1996), while fortis stops remain voiceless. Another unique difference arising in word-medial position is closure duration; fortis stops manifest a much longer closure duration than lenis stops (e.g. 126 vs. 45 ms in Han 1996). The redistribution of available acoustic cues indeed

gives rise to significant consequences in perception. In word-medial position, the once-reliable cue for the contrast, VOT, is demoted to be quite useless (Han 1996). Rather a new temporal parameter, closure duration, leads to a near-categorical boundary shift from lenis to fortis as the duration of stop closure increases.

In this study, we set closure duration to be neutral at around 80 ms, the value reported in previous studies (Han 1996), favoring neither fortis nor lenis judgments. The significant demotion of the VOT cue in word-medial position is likely to provide the best chance for the naïve listeners to make better use of the *f0* cue for stop identification. This leads to a prediction that the group difference with respect to the use of *f0* cues may be reduced, but it remains to be seen to what extent the reduction of the dominance of VOT could compensate for the lack of particular training with *f0* cues.

Another variable added to the current study is a variation in tone contours. Among the four lexical tones in Mandarin Chinese, two tonal categories were chosen for acoustic manipulation of *f0*, namely high-level (X^{55}) and mid-rising (X^{35}). Low-dipping (X^{214}) and high-falling (X^{51}) tones were excluded from the study because *f0* manipulation is not as straightforward due to frequent creaky voice (e.g. Belotel-Grenie & Grenie 1994; Kuang 2017). However, it is assumed that low-dipping and high-falling tones would pattern together with mid-rising and high-level tones, respectively, considering their comparable *f0* values at vowel onset. Note that this grouping is consistent with the observations in past research (Ko 2000; Maeng & Kwon 2008). While the variation in tones is expected to strengthen the empirical validity of the experimental findings, the comparison between level and contour tones is also likely to help to uncover an interesting nature of stop perception in general. In a cross-linguistic perception study, Oglesbee (2008) manipulated *f0* in two ways, one along the *f0* and the other along the global *f0* contour. The results showed that Japanese listeners were sensitive to the *f0* contour such that a falling contour elicited more voiceless stop responses than the corresponding level contour. Korean and Spanish listeners also showed the same direction, albeit to a smaller degree. In addition to *f0*, the tone contour was expected to shed light on the effect of changes in *f0* frequency on stop perception.

2. Method

2.1 Participants

To exclude any confounding effects of differential baseline with respect to experience in tonal languages, Korean listeners were limited to Seoul speakers whose

prosody is uncontroversially non-tonal and non-stressed, but only phrase-accented (Jun 1993). Two independent groups of Seoul speakers were recruited for two experiments. For the word-initial stop identification, seventeen graduate students of the Chinese department at Seoul National University (14 female and 3 male) participated as experienced learners. This group of speakers has specialized in the Chinese language in college and continued to learn Mandarin Chinese in graduate school. Most of them have lived in mainland China as long as two years and have passed old HSK (Chinese Proficiency Test, level 1 being the lowest and level 12 being the highest) level 6. According to the HSK Institute, those who passed the old HSK level 6 can read Chinese newspapers and magazines and appreciate Chinese movies and plays. Learners at this stage can also give a lengthy speech in Chinese. Another seventeen Seoul speakers (16 female and 1 male) were recruited for the naïve listener group. They were undergraduate students of Seoul National University who came from various specializations and had no previous experience in Mandarin or other tonal languages.

An equivalent control of participants was made for the second experiment examining the perception of word-medial stops. Seventeen graduate students of the Chinese department (15 female and 2 male) participated as experienced learners, and another seventeen undergraduate students (14 female and 3 male) participated as naïve listeners. None of them took part in the first experiment. All subjects participated in the experiment voluntarily. None of the participants had a history of hearing impairment, nor reported any present hearing disorders.

2.2 Stimuli

Disyllabic words in Mandarin Chinese were selected for the base stimuli. Real Mandarin words were used for the sake of the fluency of the talker, but the stimuli words used in this experiment are relatively rare in colloquial Mandarin, and few experienced listeners reported that they could recognize word meaning during the experiment. Even if some subconscious semantic retrieval was available in some cases, the retrieval of lexical items is likely to hinder the leaners from perceiving fine-grained phonetic details in the stimuli and can only neutralize the differences between the learners and naïve listeners. Therefore, the significant group differences, if any, will provide even stronger evidence for the effect of learning of a tone language on stop identification.

For the word-initial stop identification, the first syllable was the target syllable, and for the word-medial stop identification, the second one was the target syllable. In both cases, the vowels following the target stop sounds were fixed with [a] or [aɪ]. The preceding vowels for the word-medial stops varied between [a] and [i], but they were balanced across level and rising tone conditions. The tar-

get syllables carried either a level or rising tone and the non-target syllables were fixed with a high-level tone (X^{55}) for both experiments. The stop consonants of the target syllable varied between bilabial and alveolar stop consonants. Velar consonants were not included as targets due to an accidental gap of /ka^{35}/ in the Mandarin syllable inventory. In addition to the four target words (2 places * 2 tone contours), seven filler items with varying place of articulation for stops and tone contours were included for each word-initial and word-medial stimuli list.

For the recordings of the stimuli, one female Mandarin speaker in her twenties was recruited. She was born and raised in Beijing and has been living in Korea for about two years at the time of recording. The recording was made in a sound-attenuated booth in the Phonetics Lab in the Department of Linguistics at Seoul National University. The list of stimuli was repeated six times in random order and presented to the talker on sheets of paper (11 tokens * 2 word position * 6 repetitions, 132 tokens in total). The talker was asked to read each target word at a normal speech rate. The recording was made using a Tascam DV-RA1000 recorder and digitized at 22,050 Hz sampling rate in 16 bit using CoolEdit 2000. The recordings were normalized at 70 dB for RMS intensity using a Praat script before further analyses.

The target syllables were first annotated for the initial consonant and the following vowel. The start and end of the noise period of a stop sound were labeled at the zero-crossing and the duration between these two time points was considered the VOT value. The start and end of voicing were labeled and the *f0* values at the two time points were measured. Both VOT and *f0* values were extracted automatically using Praat scripts. All stop sounds under investigation had short but positive VOTs (mean values less than 10 ms) regardless of the position in a word. Among the six tokens recorded, one representative token with a clear and natural pronunciation of each stimulus was chosen to be used for the following perception study. The phonetic properties of the selected tokens are presented in Table 1.

Table 1. Experimental stimuli and their acoustic characteristics

	Tone	Word	VOT (ms)	*f0* contour (Hz)	Vowel duration (ms)
Initial stops	Level	**pa^{55}**.tɕ^{h}æn^{55} 'eight thousand'	6.0	275–268	139
		ta^{55}.tɕ^{h}aŋ^{55} 'respond'	8.0	282–282	144
	Rising	**pa^{35}**.tɕæn^{55} 'outstanding'	7.0	208–245	162
		ta^{35}.k^{w}an^{55} 'optimistic'	7.9	198–227	184
Medial stops	Level	tɕ^{h}i^{55}.**pa^{55}** 'seven-eight'	7.9	294–287	265
		fa^{55}.**tai^{55}** 'stare blankly'	7.7	270–257	301
	Rising	hei^{55}.**pai^{35}** 'black and white'	9.5	225–245	420
		fa^{55}.**ta^{35}** 'developed'	9.4	190–204	323

Each selected token was manipulated to create a set of words whose *f0* was on a continuum of eleven steps. This was achieved by lowering or raising the overall *f0* contour by 10 Hz at a time using the PSOLA algorithm in Praat (Boersma & Weenink 2017). The *f0* of the vowel onset was maintained to fall between 200 Hz and 300 Hz; for example, [pa] in [tɕ^{h}i^{55}.**pa**^{55}] originally started off with 290 Hz and the entire *f0* contour of the syllable was lowered or raised to create multiple tokens with *f0* from 200 Hz to 300 Hz. This procedure is illustrated in Figure 2. The particular range of *f0* was chosen because *f0* manipulation did not result in any unnatural or odd sounding signals, while it gave rise to a sufficiently gradient continuum of *f0*. In total, forty-four target stimuli (4 original tokens * 11 step manipulation) were created for each experiment. In addition, the same number of filler words with the same *f0* manipulation was added.

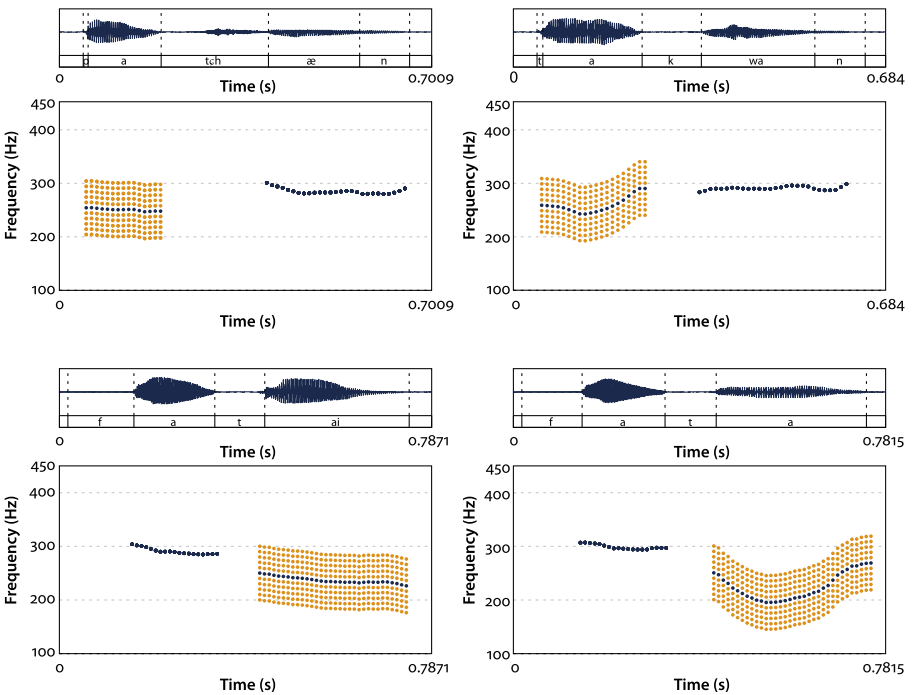

Figure 2. Example of *f0* manipulation of the stimuli for the first syllable [**pa**.tɕ^{h}æn^{55}] (level) vs. [**ta**.k^{w}an^{55}] (rising) and for the second syllable [tɕ^{h}i^{55}.**pa**] vs. [hei^{55}.**pai**]. The *f0* of the target syllable was gradually lowered or raised by 10 Hz at each time. Original *f0* contours (navy blue) and manipulated *f0* contours (orange)

One advantage of this particular method of *f0* manipulation is that the contribution of tone contours to stop perception can be assessed independently from that of *f0*. In previous research, offset *f0* was often fixed while *f0* was manipulated to be on a continuum. In contrast, in the present study, overall *f0* contour was shifted along with *f0* variation so that the intrinsic tone contour could be retained. In this way, *f0* and tone contours can enter statistical models as orthogonal factors, enabling an assessment of the independent role of each variable in stop identification.

As for the word-medial stops, the closure duration was set to be 80 ms for all stimulus items by splicing the silent period. The particular closure duration was chosen based on the previous findings. Han (1996), among others, showed that a categorical boundary shift occurred at 80 ms of closure duration, namely fortis judgments for longer closer duration vs. lenis judgments for shorter closure duration, all else being equal. At this closure duration, therefore, the stimuli should not trigger biases toward either fortis or lenis, and the sole effects of experimental variables are likely to arise clearly.

2.3 Procedure

The perception study was run in DMDX software (Forster & Forster 2003), the script of which was mounted on a Toshiba Satellite Pro laptop computer. The experiment was conducted in quiet rooms, mostly in dormitory rooms or in unoccupied classrooms. Participants were given verbal instructions about the overall procedure of the experiment in detail prior to the experiment and had five practice items before the beginning of the test trials. During instruction, they were told that they would hear non-sense words in a foreign language that they may or may not know. In the experiment, two words in printed Korean orthography appeared on the screen, one on the left-hand side and the other on the right-hand side, immediately after a stimulus was played. As in typical two-forced choice identification tasks, the participants had to make either fortis or lenis choice for each stimulus (e.g. "파다이" (lenis) vs. "파따이" (fortis) for "fa.tai"). Participants were instructed to press the key "A" for the choice on the left-hand side and "L" for the choice on the right-hand side. They were told to use both hands and place their comfortable fingers on each key button so that they could press the intended button correctly and quickly. Participants were encouraged to feel free to make responses as there were no correct answers, and were told to press either button even when they were not certain about the choice.

The presentation of Korean orthography was made possible using JPG files, and the randomization of the order of the lenis and fortis choice was made prior to the script being submitted to DMDX. A pair of common earbud earphones was

used for all participants. Participants were given 2.5 seconds to make a response, and if no response was made within this period, the program was designed to proceed to the next trial. The whole experiment took about fifteen minutes, including registration, for most participants.

The effect of *f0* on stop identification was analyzed using mixed-effects logistic regression models. Logistic regression analyses were employed since the dependent variable had two categories, namely fortis (coded as "1") vs. lenis (coded as "0"). Two separate statistical models were run, one for the word-initial stops and another for the word-medial stops. In each model, fixed effects included *f0* (centered by z-score transformation), learning experience (2 levels: learners vs. naïve), and contour (2 levels: level vs. rising). In addition to the main effects, interaction terms were also included between *f0* step and learning experience and between contour and learning experience. Recall that the main goal of the present study is to examine whether experience in Mandarin Chinese would lead to a significant increase in listeners' weighting on the *f0* cue for stop identification. A significant interaction between *f0* and learning would indicate that learning experience has triggered meaningful changes in cue weighting strategies, while a non-significant interaction would provide no evidence for the effects of learning on such changes. The interaction between contour and learning was also included to examine whether the effect of tone contour is more pronounced for the learners. As for the random effects, subject was included as random slopes for both models. Initially, an attempt was made to keep the maximal random effect models following suggestions in the literature (Barr et al. 2013), but fully crossed models did not converge and, therefore, interaction terms were dropped from the statistical models. All statistical analyses were implemented using the package *lme4* (Bates et al. 2014) available in R software (R Development Core Team 2016).

3. Result

3.1 Word-initial stop identification

Figure 3 plots predicted logit curves and actual mean data for word-initial stop identification as a function of *f0* and learning experience. The results were broken down by tone contour. One tendency that arises clearly is that experienced learners showed a near-categorical perception of the stops with a relatively abrupt shift at intermediate *f0* frequencies (approximately 240–260 Hz) for both contour conditions; stimuli with lower *f0* were identified predominantly as lenis stops, while stimuli with higher *f0* as fortis stops. While a steady and gradual increase of fortis judgments is observed for the naïve listeners as well, the pattern is far from being categorical. Rather, participants most often identified the stops as fortis.

Table 2 summarizes the results of the statistical model fit. The results confirmed significant main effects of both *f0* and learning experience, indicating that higher *f0* is more likely to elicit more fortis judgments and that naïve listeners are more likely to give more fortis judgments than experienced learners. In addition to the main effects, the results revealed a significant interaction between them. Experienced learners relied heavily on *f0* in the identification of stops ($\beta = 0.8706$), and this tendency decreased in magnitude in naïve listeners' stop identification ($\beta = 0.3158$, inferred from $0.8706 - 0.5548$). This provides strong empirical evidence for the greater weighting on *f0* for the stop identification by the experienced learners than by the naïve listeners, suggesting an effect of learning on perceptual cue reorganization.

Nevertheless, it is worth noting that naïve listeners did also systematically respond to *f0* in the same direction, albeit to a smaller degree. The results can be interpreted such that *f0* is a minor but legitimate perceptual cue used to identify stops for Korean listeners, in general. The short VOTs of the stimuli, however, appear to be sufficient to draw largely uniform responses as evidenced by the dominance of fortis responses by the naïve listeners. This is consistent with previous studies in which the fortis-lenis distinction is primarily cued by VOT alone (Kim 2004; Lee et al. 2013). This L1 cue-weighting strategy seems to have been transferred straightforwardly to naïve listeners' identification of Mandarin unaspirated stops.

The interaction between contour and learning experience did not reach statistical significance, but it is worth a comment as the *p*-value shows marginal significance ($p = .077$). Experienced learners tended to make a more fortis judgment when the target syllables carried level contour as evidenced by the negative coefficient value ($\beta = -0.2907$), but the effect was not robust enough to give rise to statistical significance ($p = .208$). The positive coefficient value of the interaction term ($\beta = 0.5955$) indicates that this tendency is reversed for the naïve listeners ($\beta = 0.3048$, inferred from $-0.2907 + 0.5955$). In the separate model with naïve learners as a baseline, the results again showed that the effect of contours did not reach statistical significance ($\beta = 0.3049$, $z = 1.246$, $p = .213$). Taken together, the ostensibly small interaction effect between contour and learning arises from the responses to target contours in the opposite direction by the two groups of listeners.

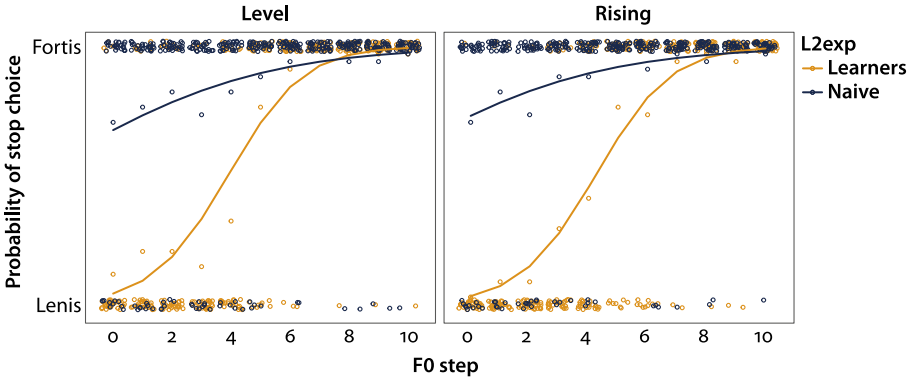

Figure 3. Predicted logit curves and actual mean values (circles) for the identification of word-initial stops for learners (orange (online)) and naïve listeners (navy blue (online)). Responses of the level contour are on the left panel, and those of the rising contour on the right panel

Table 2. Summary of the mixed-effects logistic regression model for the word-initial stop identification. Significant results are highlighted in bold

Predictor	Coefficient	*SE*	*z*-value	*p*
Intercept	**0.9741**	**0.2898**	**3.361**	**< .001**
f0	**0.8706**	**0.0608**	**14.310**	**< .001**
Learning (Learners vs Naïve)	**1.5404**	**0.4236**	**3.637**	**< .001**
Contour (Level vs. Rising)	−0.2907	0.2306	−1.260	= .208
Interaction= *f0Learning**	**−0.5548**	**0.0754**	**−7.359**	**< .001**
Interaction= Contour*Learning	0.5955	0.3362	1.771	=.077

3.2 Word-medial stop identification

Response curves and actual mean data for the word-medial stops are plotted against *f0* and learning experience in Figure 4. Most strikingly, the near-categorical perception by the experienced learners in the first experiment was no longer attested in the second experiment. While *f0* was still made use of in both contour conditions, as evidenced by the gradual increase of fortis responses along with higher *f0*, participants' responses were sharply divided by contours: fortis judgments were populated predominantly in the level contour condition and lenis judgments in the rising contour condition. Naïve listeners appeared to show some sensitivity to *f0*, but the stop identification by this group of listeners seems to be similarly influenced by the tone contour.

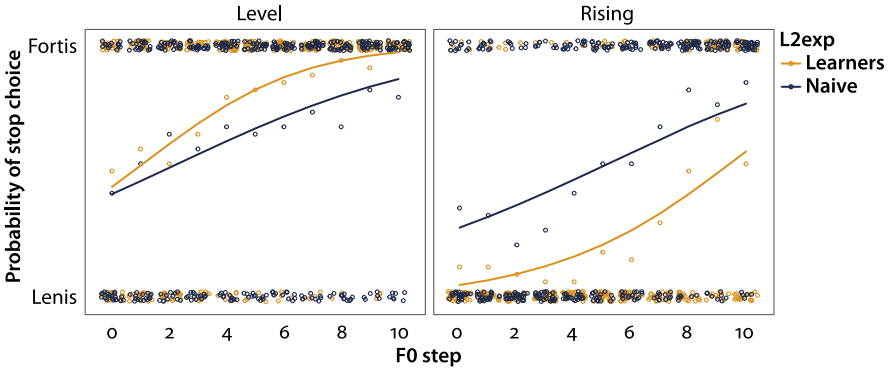

Figure 4. Predicted logit curves and actual mean values (circles) for the identification of word-medial stops for learners (orange (online)) and naïve listeners (navy blue (online)). Responses of the level contour are on the left panel, and those of the rising contour on the right panel

Table 3. Summary of the mixed-effects logistic regression model for the word-medial stop identification. Significant results are highlighted in bold

Predictor	Coefficient	*SE*	*z*-value	*p*
Intercept	1.9977	0.3678	5.432	< .001
f0	0.4470	0.0430	10.41	< .001
Learning (Learners vs. Naïve)	−1.0683	0.5010	−2.132	<.05
Contour (Level vs. Rising)	−3.7326	0.2831	−13.186	< .001
Interaction= f0 *Learning	−0.1830	0.0520	−3.519	< .001
Interaction= Contour*Learning	2.9702	0.3326	8.930	< .001

The statistical analyses summarized in Table 3 confirm the observation while revealing a slightly more nuanced nature of the data. The model fit with the level contour and experienced learners as baseline confirmed the significant main effects of *f0* and contour. The significant *f0* effect indicates that experienced learners made use of *f0* in the medial stop identification, namely high *f0* signaling fortis stops and low *f0* signaling lenis stops, similar to the findings in the first experiment. Unlike the initial stop identification, however, tone contours emerged as the most salient predictor of word-medial stops, as materialized by the high coefficient value ($\beta = -3.7326$). Level contour strongly signals fortis stops while rising contour signals lenis stops.

We did not compare participants' responses across word position in the same statistical model, but the qualitative differences between the two models are rather clear. While *f0* is no longer as salient as in the first experiment (β (initial) = 0.8706 vs. β (medial) = 0.4470) for the learners, tone contours emerged as the most infor-

mative cue for stop identification in word-medial position (β (initial) = −0.2907 vs. β (medial) = −3.7326). This sharp contrast across word position suggests that the hierarchy of perceptual cues rearranges itself according to the environment. It appears that the *f0* cue overrides other phonetic cues in word-initial position, but it is highly underweighted in word-medial position. With the demotion of the *f0* cue, other factors such as tone contours seem to emerge as a more informative perceptual cue for stop identification.

In addition to the main effect, a significant interaction was identified between *f0* and learning. The results indicate that the two groups of listeners respond differently to *f0*; experienced learners showed higher sensitivity to *f0* (β = 0.4470) than naïve listeners did (β = 0.2640, inferred from 0.4470 − 0.1830). Nevertheless, naïve listeners do also show some sensitivity to *f0* in the same direction, a pattern mirroring the findings in the word-initial stop identification. Taken together, it can be concluded that both groups of participants utilize *f0* for stop identification in both positions in a word, but the experienced learners show consistently higher sensitivity to the *f0* cue than naïve listeners.

The statistical results confirmed another significant interaction between contour and learning experience. Both groups of listeners showed negative coefficient values for contour (β (experienced) = −3.7326 vs. β (naïve) = −0.7624, inferred from −3.7326 + 2.9702), indicating that they gave more lenis responses in the rising contour condition than in the level contour condition. More importantly, the contour-learning interaction suggests that the effect of tone contours was more robust for the experienced than for the naïve listeners. This is clearly illustrated in Figure 4; for each *f0* level, the difference in mean fortis responses between the two contour conditions is always larger for the experienced than for the naïve group. It is worth noting, however, that even the naïve learners did respond to contour categories in a similar fashion, albeit to a much smaller degree.

4. General discussion

The present study investigated the perceptual patterns of Mandarin unaspirated stops by Korean listeners with or without prior experience in the tonal language Mandarin. The core finding of the experiment, especially in word-initial position, is that *f0* was the single most important perceptual cue to stop identification for the Mandarin learners. They showed near categorical perception of initial stops, in particular, whereby the stop perception shifted abruptly from lenis to fortis at intermediate *f0* values. Despite the short VOTs carried by the stimuli, low *f0* strongly signaled lenis stops while high *f0* was a strong indication of fortis stops. This pattern was sharply contrasted with that of the naïve listeners who perceived

most stimuli predominantly as fortis regardless of *f0*. The results of the present study, therefore, strongly suggest that the split in the perception of the Mandarin unaspirated stops conditioned by tone is attributed to a learning experience in the tonal second language. It follows that a significant developmental change, as summarized below, has occurred in the L1-L2 stop category mapping during the acquisition of a tonal second language by Korean listeners.

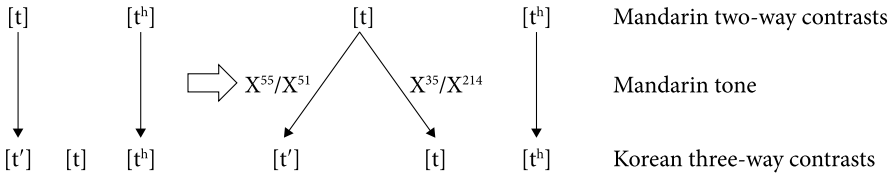

Figure 5. Tone-induced split in stop category mapping by Korean L2 learners over the course of the acquisition of Mandarin Chinese

In what follows, we discuss the implications of the findings of the current study for second language perceptual learning. In particular, we sketch the potential driving force behind the developmental change schematized above. We will then elaborate on the factors contributing to the similarities and differences in the ways in which the stops are perceived in different word positions.

4.1 One-to-two category mapping of the Mandarin stops by the learners

As reported in previous perception studies (Kim 2004; Lee et al. 2013), unambiguous VOT values take precedence over *f0* cues and provide a decisive cue to the lenis-fortis contrast for Korean listeners. Consistent with those findings, naïve listeners in the present study identified the stimuli with short VOTs most often as fortis stops. It is worth emphasizing that the stimuli used in the current study were created based on naturally produced Mandarin utterance, while stimuli used in previous studies were manipulated based on utterances of Korean words. The fact that Korean listeners show consistent identification patterns across studies despite some language-specific phonetic differences (e.g. the presence of tonal contour in Mandarin-based stimuli in the present study) once again confirms the undominated status of VOT as the most reliable phonetic cue to the word-initial fortis identification.

This heavy weighting of VOTs for fortis identification is no longer at play for the learners of Mandarin Chinese, however. The *f0* cue, instead, emerged as a strong phonetic cue, and the sharp discrepancy in VOT values between the stimuli (short VOT) and native lenis stops (long VOT) were disregarded by the learners. Why does this reorganization of phonetics cue weighting take place for the

learners of Mandarin Chinese? A simple hypothesis may attribute the change to the learners' increased sensitivity to the *f0* cues through an extensive training in tone, which may have driven a promotion of *f0* over VOT cues.

While learners must have developed a keen sensitivity to *f0* changes in the speech signal in order to learn a language where *f0* cues are used to mark lexical contrasts, empirical evidence suggests that the scenario solely based on the sensitivity to *f0* cues is insufficient to account for the perceptual change of the learners. For one thing, no change in the one-to-one mapping between aspirated stops in the two languages suggests that VOT cues are not entirely overridden by *f0* cues. If *f0* is the single most important undominated cue for the stops, Mandarin *aspirated* stops carried by a mid-rising tone (e.g. t^ha^{35}) should elicit a lenis perception while the same stops by a high-level tone (e.g. t^ha^{55}) elicit a fortis perception, leading to a perceptual split as in the case of the unaspirated stops. The lack of such evidence suggests that VOT cues still by and large play an important role in stop identification for the learners. More specifically, the Mandarin aspirated stops show long VOT values ($M = 103$ ms, averaged across different places of articulation, Rochet & Fei 1991), longer than the Korean aspirated stops ($M = 80$ ms, Lee & Jongman 2012). This is likely to make it even harder for Korean listeners to interpret the Mandarin aspirated stops as Korean lenis ($M = 65$ ms).

Another piece of counter-evidence comes from the experiments with native speakers of tone languages. In experiments testing perceptual assimilation patterns (Kim 2006; Holliday 2014), Mandarin-speaking learners of Korean showed considerable difficulties in the discrimination of Korean lenis and aspirated stops, because both stops are assimilated to a single category in Mandarin, namely aspirated stops. The salient differences in onset *f0* between lenis (low *f0*) and aspirated (high *f0*) in Korean stops are simply disregarded. Rather, their similar VOT values hinder the establishment of separate categories for Mandarin learners of Korean. Likewise, native speakers of the Gyeongsang dialect, a Korean dialect well known for its tonal properties, did not use *f0* cues in stop identification as strongly as Seoul speakers did (Lee et al. 2013). The empirical findings in the previous studies, therefore, indicate that *f0* changes in the speech signal are reserved for lexical contrasts for the speakers of tonal languages, and their high sensitivity to *f0* cues is not utilized for segmental contrasts.

With the exclusion of a few possibilities, it becomes clearer that the cue reversal between VOT and *f0* for the identification of fortis is a unique behavior of the learners, and the key to understanding the pattern should bear on the effect of learning on perceptual assimilation. As discussed in the introduction, PAM-L2 highlights the role of the acquisition of the L2 phonological system by the learners. While acoustically quite distinct, a rhotic sound in an L2 may be equated to an L1 rhotic due to similarities in their phonological behavior. As detailed in Lindau

(1985), rhotics tend to show similar phonotactic distributions (being close to the syllable nucleus), they are subject to deletion in word-final position, and they often replace each other depending on phonological environments. When learners discover the phonological distribution of the rhotics, their perception may be reshaped in a way that the corresponding rhotics may sound more similar than their actual phonetic-acoustic distance.

The tone-induced split in the perception of the Mandarin unaspirated stops might also be accounted for with reference to the learners' awareness of the L2 sound system. To elaborate, the acquisition of the phonetics and phonology of Mandarin stops may lead learners to make an implicit comparison of the Korean vs. Mandarin stop systems, and the two-way contrast in the L2 stop system may have been forced to split into a three-way contrast as in the L1. But, of course, not all foreign two-way stop contrasts are split and systematically mapped onto three-way contrasts by Korean L2 learners. Studies have reported a broad one-to-one mapping between English and Korean stops: English aspirated stops being mapped onto Korean aspirated while English voiced stops onto Korean lenis (Schmidt 1996; Park & de Jong 2008).

The particular distribution of the salient phonetic cues may be an essential key to the understanding of the category split. Unlike stops in other non-tonal languages, the stops in Mandarin carry two extreme values of onset *f0*: high (as in initial high-toned syllables X^{51} and X^{55}) or low (as in initial low-toned syllables X^{35} and X^{214}). Such bimodal distribution of onset *f0* in Mandarin may have facilitated labeling of two separate categories by the learners. Phonetic category learning based on statistical distributions of perceptual cues has shown to be highly robust for both adults and infants (Maye & Gerken 2000; Clayards et al. 2008; Maye et al. 2008). Maye et al. (2008), for example, demonstrated that listeners are sensitive to probabilistic distributions of the phonetic cues in the stimuli and come up with different generalizations over input stimuli with differential structures. In their study with English-speaking infants, a bimodal condition consisted of stops with a high occurrence of VOT values near two endpoints (i.e. –75 and 14 ms), while a unimodal condition consisted mostly of tokens with intermediate VOT values. The results showed that the bimodal group showed significantly better discrimination of novel VOT contrasts compared to the unimodal group. In the case of the Korean L2 learners of Mandarin, a continuous exposure to non-native stops with a bimodal distribution of *f0* may have facilitated separate category labeling for stops with low (lenis) vs. high (fortis) *f0*. Without the knowledge of phonetic as well as phonological properties of Mandarin stops, however, naïve listeners are likely to process speech signals at the acoustic-phonetic level and

the short-lag VOTs of Mandarin unaspirated stops may be so prominent to elicit single-category labeling.

This tone-induced split for stops holds a quite unique place among many L2 perceptual assimilation studies focusing on the cases of Single-Category Assimilation where two foreign phones assimilate to a single native phone, causing much difficulty for their discrimination. The results of the present study suggest that the knowledge of L2 phonetics and phonology is rather far-reaching than one might think and may exert strong influences on rephonologization of the L2 sound system; the existing cue-weighting strategies for L1 sounds are easily overridden by alternatives that are responsible for the ways in which L2 sounds are regrouped into new phonological categories. Sufficient empirical data of such cases need to be compiled to elaborate on the implications of PAM-L2, and original proposals with testable hypotheses in laboratory experiments is a pressing topic for future studies.

4.2 Auditory contrast effects for contour tone

The present study included the perception of stop identification in word-medial as well as in word-initial position. This addition aimed to examine whether the demoted VOT cue for the stop identification in word-medial position may lead to a reduction of the differences between the two groups of listeners. The results showed that while the reliance on *f0* cues is indeed reduced a great deal, learners still show a higher sensitivity to *f0* than naïve listeners as evidenced by the significant interaction between and *f0* and learning experience. This, once again, highlights the strong and uncontroversial effect of the learning of a tonal language on the organization of the cues for stop identification.

A rather unexpected but interesting asymmetry is that the effect of tone contours was different between initial and medial stop identification. While tone contours played a controversial role in the first experiment with initial stops, it emerged as the strongest predictor for the word-medial stops. Note that this tone contour effect was obtained independently from the effect of *f0* which exerted a consistently significant influence on stop perception across different tone contours. With varying magnitudes, both experienced and naïve listeners showed high sensitivity to tone contours in the same direction; stops with level contour elicited more fortis judgments, all else being equal, while those with rising contour elicited more lenis judgments. As indicated by the significant interaction, this tendency was much more pronounced for the experienced than for the naïve listeners. Because both tone contours were preceded by a fixed level tone in the

first syllable, the significant perceptual differences between the two contour conditions can only be attributed to the presence of the internal contour.

This asymmetry may be captured with reference to the intrinsic properties of the two contours. Specifically, a level contour lacks an intrinsic contour, whereas a rising contour involves a large change in *f0* from low to high over the entire syllable. The high *f0* in the subsequent portion of the rising contour may elicit an auditory contrast such that the low *f0* at the beginning of the vowel is perceived as even lower, which provides positive evidence for lenis stops. In contrast, the level contour is perceived rather veridically due to the lack of an intrinsic contrast, leading to more fortis judgments for the acoustically identical *f0*. The finding that experienced listeners showed the effect in a greater magnitude also makes sense, given that they are more sensitive to *f0* changes used in the speech signal. Moreover, the listeners' sensitivity to the specific tone contours has general implications for the studies of stop identification. In addition to the locally defined *f0*, the identification of stop sounds arises from the integration of the global *f0* contour as well, echoing Oglesbee (2008).

The underlying mechanism of the effect of the intrinsic tone contour is presumably analogous to the extrinsic context effects attested in lexical tone and stop perception. It is well established that tone language speakers identify a tone as *high* when preceded by a *low* tone, while the physically identical tone is perceived as *low* when preceded by a *high* tone (Moore & Jongman 1997; Wong & Diehl 2003; Francis et al. 2006). Likewise, studies have observed an asymmetry in stop perception depending on local contexts; for an ambiguous stop spectrum, listeners tend to hear more [d]s (spectrally *higher* stops) next to [ɹ] (spectrally *lower* liquid) while [g] perception (spectrally *lower* stops) is more frequent when adjacent to [l] (spectrally *higher* liquid) (Mann 1980; Lotto & Kluender 1998). The unique contribution of the current finding is that the auditory contrast in tone is further integrated to draw systematic asymmetries in stop perception.

Yet one question still remains to be answered. Why did this contrast effect hold mostly in the second experiment with the word-medial stops? Recall that a minor effect of this kind was found in the first experiment with word-initial stops. Experienced learners in the first experiment showed a trend of giving more lenis responses in the rising contour condition, but the effect was not statistically significant. This trend was, however, greatly enhanced in the second experiment with the word-medial stop identification particularly for the experienced learners. While drawing a full picture of this pattern is beyond the scope of this study, two scenarios or a combination of the two seem plausible. One possibility is that in word-initial position, phonetic cues such as *f0* (or VOT depending on the listeners' language experience) are so salient that they take precedence over other factors such as contrast effects from a rising *f0* contour. Another possibility is that in

the second or non-initial position, it is inherently easier to perceive tone contours because of the way in which speech signals unfold in time, and those in the initial or preceding syllables are likely to serve as better references.

5. Conclusion

The present study tested perceptual assimilation patterns of the Mandarin unaspirated stops by Korean listeners with or without experience in Mandarin Chinese. Compared to naïve listeners who interpreted the stimuli with short-lag VOTs mostly as fortis stops, learners showed a dominance of *f0* over VOT cues: low *f0* strongly signaled lenis stops while high *f0* signaled fortis stops. This tone-induced split in category mapping by the learners was discussed with reference to the statistical learning mechanism whereby the bimodal distribution of *f0* frequencies in Mandarin tones may have facilitated the establishment of two separate stop categories. The findings of the present study contribute a novel case to the literature in which the knowledge of L2 phonetics and phonology may be so pervasive as to drive the rephonologization of the L2 sound categories using distinct cue-weighting strategies from those in the native system. It is hoped that more empirical data will become available to better understand the role of the knowledge of L2 phonological structures in the development of second language acquisition.

Acknowledgements

I would like to thank Sean Martin for sharing his valuable insights on this work. I acknowledge reviewers of *Language and Linguistics* for their constructive comments, which led to substantial improvements to the manuscript. The manuscript has also improved much through helpful feedback from the audience at the annual meeting of the Acoustical Society of America, Seoul International Conference on Speech Sciences and NCTU Phone Brown Bag. This work was partially funded by MOST International Conference Travel Grant (106-2914-I-009-024-A1), NCTU Travel Grant, and MOST Research Grant (106-2410-H-009-049).

References

Aoyama, Katsura & Flege, James Emil & Guion, Susan G. & Akahane-Yamada, Reiko & Yamada, Tsuneo. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32(2). 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Barr, Dale J. & Levy, Roger & Scheepers, Christoph & Tily, Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas & Maechler, Martin & Bolker, Ben & Walker, Steven. 2014. lme4: Linear mixed-effects models using Eigen and S4 (R package version 1.0–6). (https://cran.r-project.org/web/packages/lme4/index.html) (Accessed 2017-04-01.)

Belotel-Grenie, Agnès & Grenie, Michel. 1994. Phonation types analysis in Standard Chinese. (Paper presented at the 3rd International Conference on Spoken Language Processing (ICSLP 94), Yokohama, 18–22 September 1994.)

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Strange, Winifred (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171–204. Baltimore: York Press.

Best, Catherine T. & Tyler, Michael D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, Ocke-Schwen & Munro, Murray J. (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 13–34. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Boersma, Paul & Weenink, David. 2017. Praat: Doing phonetics by computer (computer program). (http://www.praat.org/) (Accessed 2017-04-01.)

Bradlow, Ann R. & Pisoni, David B. & Akahane-Yamada, Reiko & Tohkura, Yoh'ichi. 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. https://doi.org/10.1121/1.418276

Chao, Kuan-Yi & Chen, Li-Mei. 2008. A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing* 13(2). 215–232.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Cho, Taehong & Jun, Sun-Ah & Ladefoged, Peter. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30(2). 193–228. https://doi.org/10.1006/jpho.2001.0153

Clayards, Meghan & Tanenhaus, Michael K. & Aslin, Richard N. & Jacobs, Robert A. 2008. Perception of speech reflects optimal use of probablistic speech cues. *Cognition* 108(3). 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Eom, Ik-sang. 1996. Cwungkwuke hankul phyokipepey kwanhan mwunceywa kaysen pangan [Issues on the Korean Romanization of Chinese and suggestions for improvement]. *Cwungkwuk Ene Yenkwu* [Journal of Chinese Linguistics in Korea] 4. 39–84.

Flege, James Emil. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Forster, Kenneth & Forster, Jonathan. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35(1). 116–124. https://doi.org/10.3758/BF03195503

Francis, Alexander L. & Baldwin, Kate & Nusbaum, Howard C. 2000. Effects of training on attention to acoustic cues. *Perception and Psychophysics* 62(8). 1668–1680. https://doi.org/10.3758/BF03212164

Francis, Alexander L. & Ciocca, Valter & Wong, Natalie King Yu & Leung, Wilson Ho Yin & Chu, Phoebe Cheuk Yan. 2006. Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America* 119(3). 1712–1726. https://doi.org/10.1121/1.2149768

Han, Jeong-Im. 1996. *The phonetics and phonology of "tense" and "plain" consonants in Korean*. Ithaca: Cornell University. (Doctoral dissertation.)

Han, Mieko S. & Weitzman, Raymond S. 1970. Acoustic features of Korean /P, T, K/, /p, t, k/ and /p^h, t^h, k^h/. *Phonetica* 22(2). 112–128. https://doi.org/10.1159/000259311

Holliday, Jeffrey J. 2014. The perceptual assimilation of Korean obstruents by native Mandarin listeners. *The Journal of the Acoustical Society of America* 135(3). 1585–1595. https://doi.org/10.1121/1.4863653

Holliday, Jeffrey J. 2016. Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica* 73(1). 33–51. https://doi.org/10.1159/000443312

Iverson, Paul & Hazan, Valerie & Bannister, Kerry. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America* 118(5). 3267–3278. https://doi.org/10.1121/1.2062307

Iverson, Paul & Kuhl, Patricia K. & Akahane-Yamada, Reiko & Diesch, Eugen & Tohkura, Yoh'ichi & Kettermann, Andreas & Siebert, Claudia. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87(1). B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody*. Columbus: Ohio State University. (Doctoral dissertation.)

Kang, Yoonjung. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45. 76–90. https://doi.org/10.1016/j.wocn.2014.03.005

Kim, Mi-Ryoung & Beddor, Patrice Speeter & Horrocks, Julie. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30(1). 77–100. https://doi.org/10.1006/jpho.2001.0152

Kim, Midam. 2004. Correlation between VOT and F0 in the perception of Korean stops and affricates. (Paper presented at the the 8th International Conference on Spoken Language Processing (INTERSPEECH-2004), Jeju Island, 4–8 October 2004.)

Kim, So Ya. 2006. Hangugeo pyeongeum gyeongeum gieume daehan jungguginui jigakjeog beomju yeongu [A study on the perception of Korean lenis, fortis, and aspirated stops by Mandarin speakers]. *Icwung Enehak* [Bilingual Research] 32. 52–79.

Kim, Yong Ok. 1987. *Tongyanghak ettehkey hal kesinka* [How to do oriental studies]. Seoul: Tongnamu.

Ko, Mi-Sook. 2000. Hancwung pwuncelumsowa chopwuncelumsowauy kwankyey yenkwu [A study on the relationship between segments and suprasegmental features]. *Cwungkwuk Ene Yenkwu* [Journal of Chinese Linguistics in Korea] 12. 235–256.

Kuang, Jianjing. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society America* 142(3). 1693–1706. https://doi.org/10.1121/1.5003649

Ladefoged, Peter & Maddieson, Ian. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Lee, Hyunjung & Politzer-Ahles, Stephen & Jongman, Allard. 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). 117–32. https://doi.org/10.1016/j.wocn.2012.12.002

Levy, Erika S. & Strange, Winifred. 2008. Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36(1). 141–157. https://doi.org/10.1016/j.wocn.2007.03.001

Lindau, Mona. 1985. The story of /r/. In Fromkin, Victoria A. (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 157–168. Orlando: Academic Press.

Lisker, Leigh & Abramson, Arthur S. 1964. A cross-language study of voicing in initial stop: Acoustic measurements. *Word* 20. 384–422. https://doi.org/10.1080/00437956.1964.11659830

Liu, Huei-Mei & Tseng, Chin-Hsing & Tsao, Feng-Ming. 2000. Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clinical Linguistics and Phonetics* 14(6). 447–464. https://doi.org/10.1080/026992000415877

Lotto, Andrew J. & Kluender, Keith R. 1998. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics* 60(4). 602–619. https://doi.org/10.3758/BF03206049

Maeng, Joooeck & Kwon, Yongsil. 2008. Hankwukin haksupcauy cwungkwuke phayelumey tayhan cikakyangsang yenkwu [A study on the perception of Mandarin stops by Korean learners]. *Cwungkwukhak Yenkwu* [The Journal of Chinese Studies] 44. 73–92.

Mann, Virginia A. 1980. Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics* 28(5). 407–412. https://doi.org/10.3758/BF03204884

Maye, Jessica & Gerken, LouAnn. 2000. Learning phonemes without minimal pairs. In Howell, S. Catherine & Fish, Sarah A. & Keith-Lucas, Thea (eds.), *BUCLD 24: Proceedings of the 24th annual Boston University Conference on Language Development*, 522–533. Somerville: Cascadilla Press.

Maye, Jessica & Weiss, Daniel J. & Aslin, Richard N. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science* 11(1). 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Moore, Corinne & Jongman, Allard. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102(3). 1864–1877. https://doi.org/10.1121/1.420092

Oglesbee, Eric Nathanael. 2008. *Multidimensional stop categorization in English, Spanish, Korean, Japanese, and Canadian French*. Bloomington: Indiana University. (Doctoral dissertation.)

Park, Hanyong & de Jong, Kenneth J. 2008. Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics* 36(4). 704–723. https://doi.org/10.1016/j.wocn.2008.06.002

R Development Core Team. 2016. R: A language and environment for statistical computing (computer program). (http://www.r-project.org/) (Accessed 2017-04-01.)

Rochet, Bernard L. & Fei, Yanmei. 1991. Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception. *Canadian Acoustics* 19(4). 105–106.

Schmidt, Anna Marie. 1996. Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustical Society of America* 99. 3201–3211. https://doi.org/10.1121/1.414804

Silva, David J. 1992. *The phonetics and phonology of stop lenition in Korean*. Ithaca: Cornell University. (Doctoral dissertation.)
Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23(2). 287–308. https://doi.org/10.1017/S0952675706000911
Wong, Patrick C. M. & Diehl, Randy L. 2003. Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research* 46(2). 413–421. https://doi.org/10.1044/1092-4388(2003/034)

Author's address

Sang-Im Lee-Kim
Department of Foreign Languages and Literatures
National Chiao Tung University
R412, Humanities Bldg. 3
1001 University Road
Hsinchu, 20010
Taiwan
sangimleekim@nctu.edu.tw

Publication history

Date received: 17 April 2017
Date accepted: 21 February 2018