Linguistic Research 39(3): 567-601
DOI: 10.17250/khisli.39.3.202212.007

Native and non-native perception of Mandarin level tones*

Kyungmin Lee · Ok Joo Lee*****
(Seoul National University)

Lee, Kyungmin and Ok Joo Lee. 2022. Native and non-native perception of Mandarin level tones. *Linguistic Research* 39(3): 567-601. The present study examined how native Mandarin listeners and native Korean listeners of L2 Mandarin perceive the Mandarin level tones (i.e., Tone 1 and Tone 3) and how the Mandarin tone identification interacts with the Korean lenis-fortis stop perception in native Korean listeners. Two perception experiments, namely *Mandarin Tone 1-Tone 3 Identification Experiment* and *Korean Lenis-Fortis Identification Experiment*, were conducted. Results showed that native Mandarin listeners identified the monosyllabic level tones mostly as Tone 1, whereas the level tones on the first syllable in disyllabic expressions were distinguished as either Tone 1 or Tone 3 in a categorical manner. It was also found that the proficiency of L2 Mandarin has a strong influence on non-native tone perception, in that the responses of native Korean listeners of advanced L2 Mandarin largely conformed with those of native Mandarin listeners in both monosyllabic and disyllabic conditions. By contrast, less proficient listeners' tone judgments in monosyllabic conditions considerably correlate with the pitch height change, while a tendency to categorically distinguish Tone 1 and Tone 3 emerged in disyllabic conditions, though with less well defined perceptual boundary. The present study further revealed that L2 tones play a role in the perception of L1 stops. Despite the fact that Mandarin stops resemble Korean fortis in VOT, native Korean listeners of proficient L2 Mandarin appeared to distinguish the lenis and fortis stops in a categorical fashion in the initial position of disyllabic expressions. A great deal of similarity found between the lenis-fortis identification and the Tone 1-Tone 3 identification indicates that F0 cues are facilitated for the lenis-fortis identification of native Korean listeners. Findings of this study, therefore, demonstrate the perceptual reorganization that VOT cues are overridden by F0 was particularly notable in native Korean listeners of proficient L2 Mandarin, while the perceptual cue weighting shift did not seem to occur in less proficient L2 listeners. **(Seoul National University)**

---

* This research was supported by AIIS (Artificial Intelligence Institute, Seoul National University) and the National Research Foundation of Korea (NRF) grant funded by the Korean government [NRF-2022S1A5A2A01043267]. The authors thank the anonymous reviewers' insightful comments and suggestions on this paper.

** First author
*** Corresponding author

Keywords Mandarin, Korean, L1, L2, perception, level tone, stop, pitch, VOT

1. Introduction

Mandarin has two level tones, namely, high and low tones (Tone 1 and Tone 3 hereafter). Tone 3 has a few phonetic variants in that it is realized as low dipping tone in isolation and rising tone before another Tone 3 (see Section 2.1 for the Mandarin tones). Leaving aside a long standing controversy on which variant should be treated as the underlying form of Tone 3 in the Mandarin phonology, Tone 3 is most often realized as low tone. If we understand that low tone is the base form of Tone 3, an interesting question arises as to how the two level tones of Mandarin are distinguished from each other. Given that the production and perception of tonal pitch are considerably influenced by such factors as phonological conditions (e.g., monosyllabic vs. disyllabic conditions) adjacent tones, a speaker's pitch range, segmental properties, and the tonal inventory of a native language (Lin and Wang 1984; Fox and Qi 1990; Cao and Maddieson 1992; Ren 1992; Belotel-Grenié and Grenié 1994; Lee et al. 1996; Moore and Jongman 1997; Huang and Holt 2009; Lee et al. 2009; Cao 2010 a, b; Huang and Johnson 2010; So and Best 2010; Yang 2011, 2015; Peng et al. 2012; Zhang et al. 2012; Kuang 2013 a, b; Gao and Hallé 2015, 2017; Sjerps et al. 2018; Shi 2019), of particular interest are whether the Mandarin level tones that vary in F0 continuum are categorically identified when only pitch height cues with no contour cues are available to listeners, and how the pitch heights of the adjacent level tone affect the tone identification.

Another intriguing question pertains to perceptual differences between native listeners and non-native listeners. In particular, little is understood regarding how L2 Mandarin listeners perceive level tones of differing pitch heights when pitch height cues are also used to facilitate segmental identification in their native language (L1 hereafter). As PAM (Perceptual Assimilation Model, Best 1995; Best et al. 2001; Best and Tyler 2007) would predict, an exposure to L2 Mandarin may reshape the interpretation of tones and stops; unaspirated stops of a lower pitch may be interpreted more as lenis stops by native Korean listeners of L2 Mandarin while native Korean listeners with little exposure to Mandarin tend to associate unaspirated stops of Mandarin with fortis stops of Korean (Ko 2000; Maeng and Kwon 2008; Lee-Kim 2020). What still remains unexamined is whether native Korean listeners of L2 Mandarin develop a categorical perception in level tones and how their tone perception interacts with stop identification. If the knowledge of L2 phonology plays a significant role, we may expect F0 cues to override VOT cues in native Korean listeners' Mandarin stop identification.

The purpose of the present study is two-fold. First, it aims to provide experimental evidence regarding the questions of (1) how tones of level pitch are perceived with regard to phonological conditions (i.e., monosyllabic vs. disyllabic conditions), and (2) how native and non-native listeners differ and what role the L2 proficiency plays in perception. An examination of the perceptual difference between monosyllabic conditions and disyllabic conditions will help to distinguish the role of the pitch height cues of the tone in question from that of the following tone.[1] It also attempts to understand (1) how pitch height cues are simultaneously used in distinguishing stop consonants in L1 and perceiving level tones in L2, and (2) how phonological cue shifting takes place in non-native listeners' L1 or L2 perception. Two perception experiments were conducted for the study: *Mandarin Tone 1-Tone 3 Identification Experiment* and *Korean Lenis-Fortis Stop Identification Experiment*. Both native Mandarin listeners and native Korean listeners of L2 Mandarin participated in the Mandarin tone identification experiment, while the Korean stop identification experiment was only for native Korean listeners whose L2 Mandarin proficiency ranged from beginning to advanced levels. Given that a lenis and fortis contrast of stops in Korean is made primarily by pitch as well as VOT (voice onset time) (see Section 2.2 for the Korean stops), a close examination on the perceptual roles of pitch cues in native Korean listeners can shed light on non-native listeners' tone perception and the perceptual interaction between L1 consonants and L2 tones.

This paper is structured as follows. Section 2 gives a brief overview of the Mandarin tones with the following introduction of the Korean stops of phonation contrast. Section 3 introduces two identification experiments conducted for the study. Section 4 presents the results, and the discussion and conclusion are given in Section 5 and Section 6, respectively.

---

1 The effect of pitch height cues of the preceding level tone cannot be examined since a lower pitch on the preceding syllable will violate a phonetic constraint that two consecutive low level tones do not occur in Mandarin.

2. Background

2.1 Mandarin tones

It is widely known that Mandarin has four tones, which are labeled as Tone 1 through Tone 4 for ease of reference. The pitch contours are often described as high level, high rising, low dipping, and high falling, respectively. Table 1 presents a set of words with identical segments and differing tones in Mandarin. The pitch value is based on Chao's five-number scale, in which '1' indicates the lowest pitch and '5' indicates the highest pitch (Chao 1930, 1956, 1968: 26).

Table 1. Mandarin tones

tone number	pitch description	pitch value	example
1	high level	55	bā [pa]$_{55}$ '8, eight' 八
2	high rising	35	bá [pa]$_{35}$ 'to pull up' 拔
3	low dipping	214	bǎ [pa]$_{214}$ 'to hold' 把
4	high falling	51	bà [pa]$_{51}$ 'father' 爸

Among the four tones, Tone 3 has a few phonetic realizations. It surfaces as low tone or the so-called 'half Tone 3' when followed by Tones 1, 2 and 4, while it is changed to Tone 2 when followed by another Tone 3.[2] It can be produced as dipping tone only when it occurs in isolation or in the phrase-final position. While dipping tone was traditionally regarded as the underlying form of Tone 3 (Chao et al. 1948; Chao 1968; Cheng 1973), low tone that helps to economize articulatory effort through contour reduction has a significantly wider distribution in lexicon as well as natural speech and is the most productive variant of Tone 3 (Chen 1984; Duanmu 2007; Zhang 2010; Fu and Lee 2022). Considering that a small pitch fall on onset can be attributed to the time needed for pitch drop to the lowest level and is likely to be perceptually ignored (Maddieson 1978; Yip 2002), the base form of Tone 3 can be posited as low level tone (Chao 1932, 1933; Wang 1989; Yip 2002; Lin 2007; Shi and Ran 2011; Shi 2019).

2 Experimental studies have found that the Tone 2 derived from Tone 3 is slightly lower in pitch than the original Tone 2. However, native speakers do not perceive the differences between the two types of Tone 2 (Wang and Li 1967; Zee 1980a; Peng 2000; Yuan and Chen 2014).

2.2 Lenis-fortis contrast of Korean stops

Korean has a three-way contrast of stops: lenis (/tal/ ‘moon’ 달), fortis (/t*al/ ‘daughter’ 딸), and aspirated (/tʰal/ ‘mask’ 탈). Among several acoustic cues such as VOT, the durations of stop closure and aspiration, the durations of the following and preceding vowels, the intensity of the release burst, and the F0 of the following vowels, VOT and F0 have been found to be the primary cues in distinguishing three types of stops in the phrase-initial position. VOTs are shortest for fortis stops, intermediate for lenis stops, and longest for aspirated stops. F0 also plays a crucial role in that F0 manifests lowest in lenis stops and considerably higher in fortis and aspirated stops (e.g., Han 1996; M.-R. Kim 2000, 2014; Cho et al. 2002; Ko 2003; M. Kim 2004; Jeong 2010; Jang 2012; Kang 2014). The lenis-fortis contrast is primarily cued by VOT and F0 in production while F0 functions as the most important perceptual cue, followed by VOT (Schertz et al. 2015). Due to the merger of VOT between lenis and aspirated stops, particularly among younger speakers, F0 became the most reliable dimension that distinguishes lenis and aspirated stops (Silva 2006; M.-R. Kim 2014).

On the other hand, stops in Mandarin only have a two-way distinction: unaspirated (/ta/ dǎ ‘to hit’ 打) and aspirated (/tʰa/ tǎ ‘tower’ 塔). Unaspirated stops have considerably shorter VOTs than aspirated stops also in Mandarin, while Tones 2 and 3 syllables tend to have longer VOTs than Tones 1 and 4 syllables (Liu et al. 2008; Peng et al. 2009). Overall, Mandarin unaspirated stops resemble Korean fortis stops in VOT, and the VOT difference between the two types of stops in Mandarin is fairly comparable with that between fortis and aspirated in Korean (Joh 1996; Maeng and Kwon 2007 a, b; Shi 2019). The VOT similarity between Mandarin unspirated stops and Korean fortis stops is also supported by Lee-Kim (2020) that reports naive native Korean listeners tend to judge Mandarin unaspirated stops as fortis in Korean.

Learning experience of L2 Mandarin, however, seems to influence native Korean listeners’ stop perception. When native Korean listeners of L2 Mandarin listen to Mandarin unaspirated stops, they tend to judge syllables of Tones 2 and 3 as lenis but those of Tones 1 and 4 as fortis (Ko 2000; Maeng and Kwon 2008; Zhang and Kim 2019; Lee-Kim 2020; Lee and Kim 2022). Moreover, they tend to interpret Mandarin unaspirated stops as fortis as the onset F0 increases with no VOT changes. We can then conjecture that the exposure to L2 Mandarin may replace VOT cues with F0 cues in perception of Korean stop sounds, while the relation between L2 proficiency and cue weighting shift still remains to be further investigated.

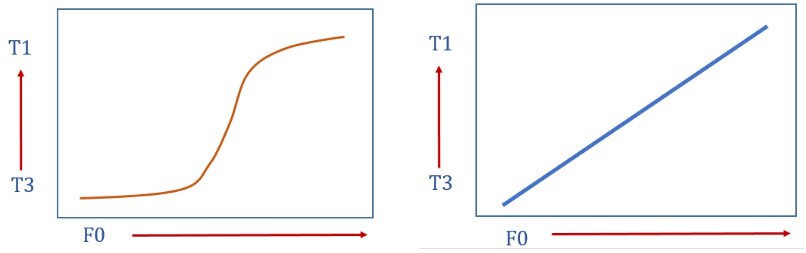
3. Research questions and experimental design

3.1 Research questions and hypotheses

The specific research questions that the present study aims to provide answers to are as follows:

1. Do native Mandarin listeners categorically identify two types of level tones that vary on continuum in pitch height?
2. Are there any perceptual differences between native and non-native listeners? Would native Korean listeners also identify the Mandarin level tones in a categorical manner? What is the role of L2 proficiency in tone perception?
3. How does the condition in which level tone appears affect listeners' perception? Do native and/or non-native listeners perceive level tones differently depending on whether they appear in isolation versus in disyllabic condition?
4. How does native Korean listeners' L2 tone perception interact with their L1 lenis-fortis stop perception? Does pitch change affect lenis-fortis judgements with little regard to VOTs? Would the perceptual interaction between tones and stops in native Korean listeners change depending on the proficiency of L2 Mandarin?

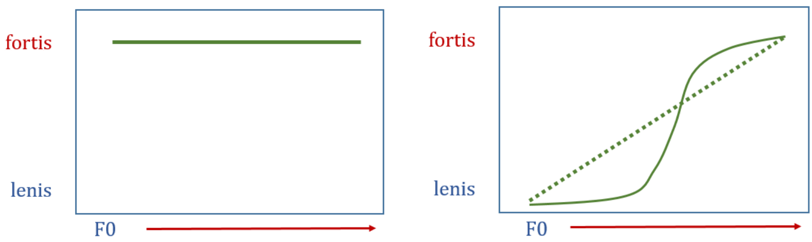
If native Mandarin listeners identify level tones in a categorical manner whereas native Korean listeners do not, the native and non-native Tone 1 and Tone 3 distinction patterns may be schematized as in Figures 1(a) and 1(b), respectively. Figure 1(a) represents an S-shape pattern with a rather quick categorical shift, and Figure 1(b) shows a pattern that the number of Tone 1 responses grows as F0 increases. It is also presumable that native Korean listeners of higher L2 Mandarin proficiency may yield more native like patterns.



(a) (b)

Figure 1. Schematized categorical and non-categorical perceptions of level tones

If F0 or tone change has little impact on native Korean listeners' stop perception, the lenis-fortis distinction characteristics may mirror the pattern schematized in Figure 2(a) since Mandarin unaspirated stops are compatible with Korean fortis stops in VOT. By contrast, if tone perception influences native Korean listeners' stop perception, the lenis-fortis identification patterns may resemble the schematizations in Figure 2(b). In the figure, the dotted line shows that fortis identification may increase as F0 becomes higher, yet with no categorical perception formed. The solid line represents a categorical identification pattern in which F0 cues override VOT cues in stop perception. It can also be conjectured that the perception characteristics may differ depending on L2 Mandarin proficiency.



(a) (b)

Figure 2. Schematized perceptions of lenis and fortis stops

3.2 Experimental design

3.2.1 Stimuli

The stimuli used in the experiments were naturally produced words with synthetic pitch contours manipulated for each of them. The words consist of twelve pairs of monosyllabic words and nine pairs of disyllabic words. In each pair, the monosyllabic word and the first syllable of the disyllabic word can be either Tone 1 or Tone 3 as both tones make naturally occurring words in Mandarin. The second syllables of all disyllabic words were also level tone, namely, Tone 1.[3] The monosyllabic words were comprised of [p], [t], [k] onsets and [a], [ai], [ɑo], [an] rimes, and the first syllables of the disyllabic words were comprised of [p], [t], [k] onsets and [ai], [ɑo], [an], [wan] rimes. Table 2 presents the list of the words used to create the stimuli.

Table 2. List of words used to create stimuli

monosyllabic words	‘ba’ [pa] 八, 把	‘da’ [ta] 搭, 打	‘ga’ [ka] 咖, 嘎
	‘bai’ [pai] 掰, 摆	‘dai’ [tai] 呆, 歹	‘gai’ [kai] 该, 改
	‘bao’ [pɑo] 包, 保	‘dao’ [tɑo] 刀, 倒	‘gao’ [kɑo] 高, 搞
	‘ban’ [pan] 班, 板	‘dan’ [tan] 单, 胆	‘gan’ [kan] 干, 感
disyllabic words	‘baikai’ [pai.kʰai] 掰开, 摆开	‘daichi’ [tai.tʂʰʅ] 呆痴, 歹吃	‘gaitian’ [kai.tʰjɛn] 盖天, 改天
	‘baogang’ [pɑo.kɑŋ] 包钢, 宝钢	‘daocha’ [tɑo.tʂʰa] 刀叉, 倒插	‘gaotong’ [kɑo.tʰuŋ] 高通, 搞通
	‘banche’ [pan.tʂʰɤ] 班车, 板车	‘duanzhuang’ [twan.tʂwɑŋ] 端庄, 短装	‘guanjia’ [kwan.tɕja] 官家, 管家

The words in Table 2 were produced by a female native speaker of Mandarin. In order to ensure that Tone 3 in monosyllabic condition was produced as low tone, it was uttered in the phrase followed by a neutral-toned syllable ‘de’ (particle 的). The speaker’s average F0 values for Tone 1 and Tone 3 were approximately 240Hz and 120Hz, respectively. In order to avoid a possible influence of segmental quality on tone perception, we counterbalanced the stimuli by synthesizing the half from the words produced as Tone 1 and the other half from the words produced as Tone 3. Each monosyllabic word and the first syllable of each disyllabic word were manipulated to

3 Tone 3 was not used for the second syllable since it triggers Tone 3 sandhi that changes the preceding Tone 3 to Tone 2.

vary in F0; the pitch was decreased by 10Hz on a 12-step continuum from the highest F0 from Tone 1 and increased by 10Hz on a 12-step continuum from the lowest F0 from Tone 3. Thirteen synthetic sounds were created for each word, and a total of 273 tokens were designed for the experiments (i.e., 13 sounds of differing pitch x 21 words = 273). Figure 3 schematizes the pitch synthesization for stimuli.

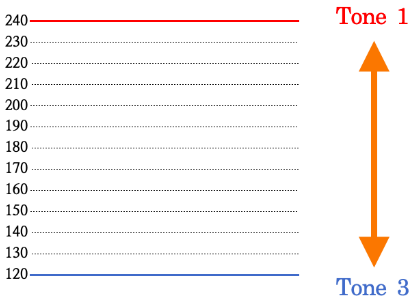


Figure 3. F0 synthesization for stimuli

3.2.2 Participants

Two groups of listeners participated in the experiments: native Mandarin listeners and native Korean listeners. Twenty three native Mandarin listeners (2 males and 21 females; 19 in their 20's, 4 in their 30's, mean age=27.91, SD=3.17) participated in Mandarin Tone Identification Experiment, and seventy native Korean listeners, namely, Korean learners of L2 Mandarin (20 males and 50 females; 68 in their 20's, 2 in their 30's, age=24, SD=2.37) participated in both Mandarin Tone Identification Experiment and Korean Stop Identification Experiment. Native Korean listeners were grouped into three according to their L2 Mandarin fluency: 24 beginning L2 listeners, 23 intermediate L2 listeners, and 23 advanced L2 listeners. The beginning L2 listeners had studied Mandarin for less than one year with no experience of staying in China, while the intermediate and advanced L2 listeners had three to five years of learning experience and more than five years of learning experience, respectively. No beginning L2 listeners have acquired a higher level score in New HSK (*Hanyu shuiping kaoshi*), a standardized proficiency test of Chinese level. Twenty one out of 23 advanced L2 listeners achieved a score at the highest level (i.e., Level 6) in New HSK or have stayed in China for more than 5 years, and the rest two were senior students who started studying Chinese at high school and were majoring in Chinese at university. The intermediate L2 listeners either have not been to China or have stayed for no longer than 18 months, while their New HSK scores range from none (i.e., no test score) to a higher level.

All participants were undergraduates or graduates at universities in Korea and were paid for their participation. They gave informed consent in compliance with a protocol approved by the IRB of Seoul National University in Korea. No participant reported speech or hearing impairment.

3.2.3 Procedure

Ninety three subjects individually participated in the on-line perception experiments run on the LabVanced program. Both Mandarin Tone Identification Experiment and Korean Stop Identification Experiment used a forced-choice labeling paradigm in which participants heard each stimuli in isolation and were asked to choose one from two choices provided on the screen. In Mandarin Tone Identification Experiment, native Mandarin listeners and native Korean listeners were instructed to click on either 'Tone 1' or 'Tone 3' after listening to each audio stimuli. They were told to identify only the first syllable when the heard stimulus was a disyllabic word. Native Korean listeners also participated in Korean Stop Identification Experiment in which they were asked to judge whether each audio stimuli sounded more similar to 'lenis' or 'fortis'. They were instructed to match only the first syllable to the Korean lenis or fortis stop, when the heard stimulus was a disyllabic word. The lenis-initial syllable and the corresponding fortis-initial syllable were given in the Korean orthography *Hangul* (e.g., '가' and '까', '바이카이' or '빠이카이').

A brief introduction regarding the Mandarin tones followed by two practice sessions was given for each experiment in order for participants to familiarize themselves with the tasks. Participants were informed that 'Tone 3' in the experiment refers to the so-called 'Half Tone 3' as in the first syllables of the words jiǎndān 'simple' 简单, jiǎnzhí 'simply' 简直, and jiǎnyào 'concise' 简要. The first practice session presented five pairs of monosyllabic stimuli using a syllable 'du [tu]' with differing pitch heights, the second practice session provided another five pairs of disyllabic stimuli using 'dache [ta.tʂʰɤ]' with varying pitch heights.

In Mandarin Tone Identification Experiment, participants first listened to 156 monosyllabic words separated in twelve blocks and then listened to 117 disyllabic words separated in nine blocks. Within each block, thirteen segmentally identical words with differing pitch heights were presented in a randomized order. The order of blocks was counterbalanced across participants. The same stimuli and procedure were applied in Korean Stop Identification Experiment. In total, each native Mandarin listener listened to 273 stimuli, and each native Korean listener listened to 546 stimuli (273 stimuli in Mandarin Tone Identification Experiment and 273 stimuli in Korean Stop Identification Experiment). In order to minimize fatigue, participants were allowed to have a short break between the blocks when needed. Including breaks, it took about 20-30 minutes for native Mandarin listeners to complete the task, and native Korean listeners took approximately 50-60 minutes to finish the two tasks.

3.2.4 Analysis

Participants' responses were grouped into four data sets: (1) Mandarin tone identification in monosyllabic conditions, (2) Mandarin tone identification in disyllabic conditions, (3) Korean stop identification in monosyllabic conditions, and (4) Korean stop identification in disyllabic conditions. For each set of data, Mixed Effect Logistic Regression Model and Multiple Linear Regression Model were adopted to analyze the identification patterns and reaction time, respectively, using R (R Development Core Team 2022) lme4 package (Bates et al. 2015). For the data sets (1) and (2), subject groups (i.e., native Mandarin listeners, native Korean listeners of beginning L2 Mandarin, native Korean listeners of intermediate L2 Mandarin, and native Korean listeners of advanced L2 Mandarin), pitch heights (120~240Hz at 10Hz interval), onset types ('b', 'd', 'g'), and rime types ('a', 'ai', 'ao', 'an', 'uan') were treated as fixed effects for Mixed Effect Logistic Regression Model and independent variables for Multiple Linear Regression Model, respectively. Subjects were random effects of the model for response. For the data sets (3) and (4), subject groups (i.e., native Korean listeners of beginning L2 Mandarin, native Korean listeners of intermediate L2 Mandarin, and native Korean listeners of advanced L2 Mandarin), task types (Mandarin Tone Identification and Korean Stop Identification), pitch heights (120~240Hz at 10Hz interval), onset types ('b', 'd', 'g'), and rime types ('a', 'ai', 'ao', 'an', 'uan') were treated as fixed effects for Mixed

Effect Logistic Regression Model and independent variables for Multiple Linear Regression Model, respectively. Subjects were random effects of the model for response. *Bonferroni* post-hoc analysis was performed for each fixed effect and independent variable using emmeans package (Lenth 2022).

4. Results

The present study seeks to understand the differences in tone perception between native and non-native Mandarin listeners, and the perceptual interaction between L2 tones and L1 segments in non-native listeners. The analysis results of Mandarin tone identification patterns and reaction time are presented in Section 4.1. Section 4.2 discusses the results of Korean stop identification characteristics and reaction time. In both sections, the results obtained from monosyllabic conditions precede those from disyllabic conditions.

4.1 Experiment 1: Mandarin tone identification

4.1.1 Identification

Mandarin tone identification in monosyllabic conditions

The Mandarin tone identification patterns in monosyllabic conditions are presented in Figure 4. In Figures 4 to 6, the x-axis and y-axis correspond to pitch height and tone response, respectively. As Tone 1 response was coded as ‘1’ and Tone 3 response was coded as ‘0’, higher mean plots on the y-axis indicate more Tone 1 responses. The responses of the four subject groups were distinguished by the line types: the sold line corresponds to the native Mandarin group, and the dashed, dotted, and dash-dotted lines correspond to the advanced, intermediate, and beginning L2 groups, respectively. A clear tendency was found in the figure that native Mandarin listeners identified the heard stimuli as Tone 1 regardless of pitch heights. While Tone 1 was preferred to Tone 3 also by native Korean listeners, the L2 Mandarin proficiency appeared to play an important role in identification. It is notable that the advanced L2 group’s identification was in line with that of the native Mandarin group. On the other hand, Tone 1 responses gradually increased over the course of 120Hz to 200Hz for both the intermediate and beginning L2 groups.

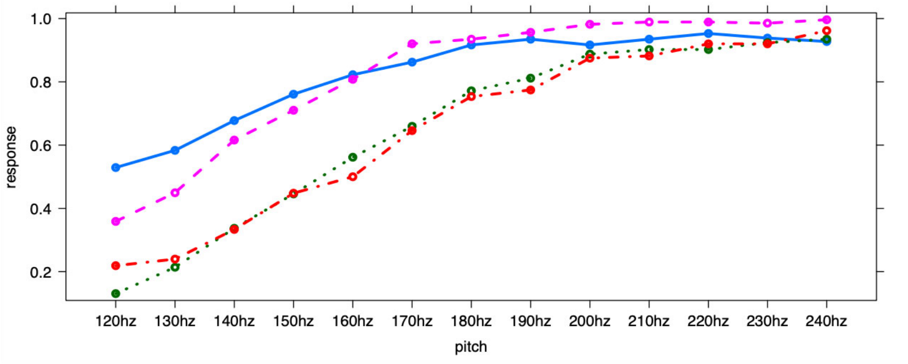


Figure 4. Mandarin tone identification in monosyllabic conditions (Mandarin, advanced L2 Mandarin, intermediate L2 Mandarin, beginning L2 Mandarin)

Mixed effect logistic regression model analyses found significant main effects of subject group (F=8.7912, p<0.001), pitch height (F=537.1286, p<0.001), onset type (F=7.798, p<0.001), and rime type (F=14.4826, p<0.001). There were also significant interaction effects of subject group*pitch (F=13.7662, p<0.001), subject group*rime (F=2.9322, p<0.01), and onset*rime (F=4.2036, p<0.001). *Bonferroni* post hoc tests confirmed no statistical group differences over the entire pitch span between the native Mandarin group and the advanced L2 group despite more Tone 3 responses yielded by the latter for the two lowest pitch heights (i.e., 120Hz and 130Hz). No significant group differences were found at all pitch heights between the intermediate and beginning L2 groups.

Table 3 presents the percentage of Tone 1 responses at each pitch height in the four subject groups. In Tables 3 and 4, the shaded areas indicate that the percentage changes were different at a significance level of 0.05. It further highlights the tendency that the native listener and advanced L2 groups tend to perceive the stimuli as Tone 1 with little influence of pitch height.

Table 3. Percentages of Tone 1 responses in monosyllabic conditions

Pitch (Hz)	120	130	140	150	160	170	180	190	200	210	220	230	240
Man	52.9%	58.3%	67.8%	76.1%	82.2%	86.2%	91.7%	93.5%	91.7%	93.5%	95.3%	93.8%	92.8%
Kor (Adv)	35.9%	44.9%	61.6%	71.0%	80.8%	92.0%	93.5%	95.7%	98.2%	98.9%	98.9%	98.6%	99.6%
Kor (Int)	13.0%	21.4%	33.7%	44.6%	56.2%	65.9%	77.2%	81.2%	88.8%	90.2%	90.2%	92.4%	93.5%
Kor (Beg)	21.9%	24.0%	33.3%	44.8%	50.0%	64.6%	75.3%	77.4%	87.5%	88.2%	92.0%	92.0%	96.2%

Mandarin tone identification in disyllabic conditions

Figure 5 shows the Mandarin tone identification patterns in disyllabic conditions. It can be immediately seen that the different tone identification patterns emerged in disyllabic conditions (i.e., the initial syllables of disyllabic expressions). Two salient perceptual differences between monosyllabic conditions and disyllabic conditions were observed. First, Tone 3 judgments were dominant for the stimuli over the lower half pitch span (i.e., 120 Hz to 180 Hz) in all four subject groups. Second, a rather abrupt change from Tone 3 response to Tone 1 response was found. A quick shift in tone identification is particularly noticeable in the native Mandarin and advanced L2 groups. This indicates that the tones were perceived in a categorical manner by both the native and non-native groups, while the perceptual boundary appeared to be relatively gradient in the less proficient L2 groups (i.e., the intermediate and beginning L2 groups).

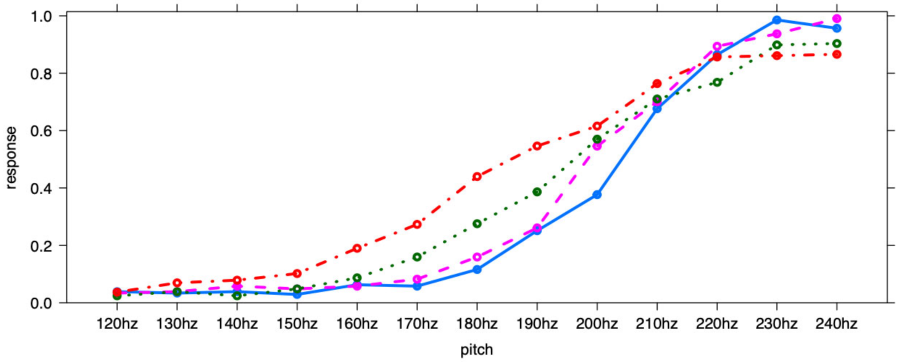


Figure 5. Mandarin tone identification in disyllabic conditions
(—— Mandarin, – – advanced L2 Mandarin, · · · · intermediate L2 Mandarin, · – · beginning L2 Mandarin)

Mixed effect logistic regression model analyses revealed that there were significant main effects of pitch height (F=1065.6851, p<0.001) and rime type (F=13.9052, p<0.001). Significant interaction effects of subject group*pitch (F=9.7094, p<0.001), subject group*onset (F=3.5171, p<0.01), pitch*rime (F=2.0178, p<0.01), and onset*rime (F=8.3951, p<0.001) were also found. *Bonferroni* post hoc tests found statistical differences neither between the native Mandarin group and the advanced L2 group nor between the intermediate and beginning L2 groups. The intermediate and beginning L2 groups were different from the native Mandarin group at the pitch of 200Hz and over the range from 170Hz to 200Hz, respectively. This confirms that the perceptual boundaries were less well defined in the less proficient L2 groups.

Table 4 shows the percentage of T1 responses at each pitch height in the four subject groups. It demonstrates that the perceptual boundary shifted to a higher pitch in disyllabic conditions with markedly less T1 responses in all four groups.

Table 4. Percentages of Tone 1 responses in disyllabic conditions

Pitch (Hz)	120	130	140	150	160	170	180	190	200	210	220	230	240
Man	3.9%	3.4%	3.9%	2.9%	6.3%	5.8%	11.6%	25.1%	37.7%	67.6%	86.5%	98.6%	95.7%
Kor (Adv)	3.4%	3.9%	5.8%	4.8%	5.8%	8.2%	15.9%	26.1%	54.6%	70.1%	89.4%	93.7%	99.0%
Kor (Int)	2.4%	3.9%	2.4%	4.8%	8.7%	15.9%	27.5%	38.7%	57.0%	71.0%	76.8%	89.9%	90.3%
Kor (Beg)	3.7%	6.9%	7.9%	10.2%	19.0%	27.3%	44.0%	54.6%	61.6%	76.4%	85.7%	86.1%	86.6%

In addition to the slopes, the categorical boundaries also differ in pitch across the subject groups. Figure 6 shows that the native Mandarin group's perceptual boundary was found approximately at 205Hz, whereas those of the non-native groups were located approximately at 199Hz, 196Hz, and 185Hz in the order of advanced, intermediate, and beginning proficiency levels. The L2 group's tone identification boundary approached that of the native group as the L2 proficiency level became higher. This indicates that not only the sharpness of identification boundary but also the pitch in which a categorical shift takes place was influenced by the L2 proficiency.

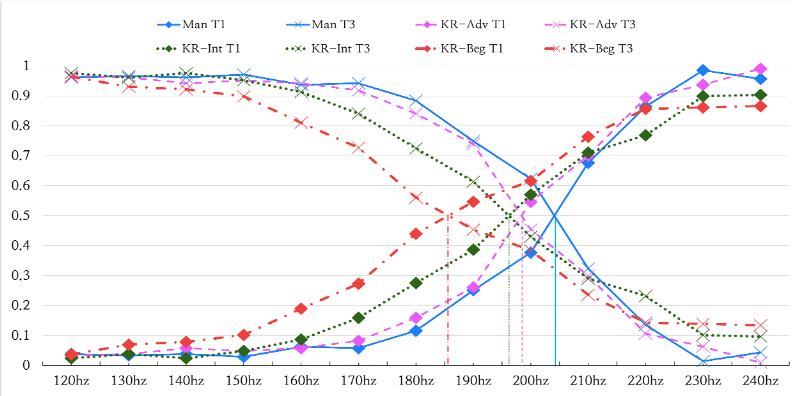
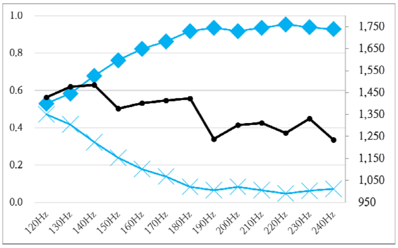


Figure 6. Mandarin tone identification boundaries in disyllabic conditions

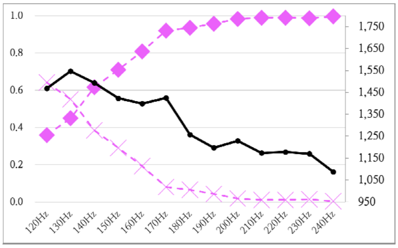
4.1.2 Reaction time

The mean reaction times for the Mandarin tone identification in monosyllabic conditions were ‘1,311 ms (advanced L2) < 1,360 ms (beginning L2) < 1,361 ms (native Mandarin) < 1,460 ms (intermediate L2)’, and those in disyllabic conditions were ‘1,344 ms (beginning L2) < 1,413 ms (native Mandarin) < 1,462 ms (advanced L2) < 1,560 ms (intermediate L2)’. Multiple linear regression model analyses found that in monosyllabic conditions, subject group (F=31.4261, p<0.001), pitch height (F=31.2929, p<0.001), and onset type (F=8.6219, p<0.001) had main effects, and interaction effects of subject group*pitch height (F=2.4635, p<0.001), subject group*onset type (F=3.0689, p<0.01), subject group*rime type (F=1.9307, p<0.05), and subject group*onset type*rime type (F=2.2384, p<0.01) were significant. In disyllabic conditions, subject group (F=55.536, p<0.001) and pitch height (F=34.7687, p<0.001) had significant main effects, and interaction effects of subject group*pitch height (F=1.6046, p<0.05) were also significant. *Bonferroni* post-hoc analysis results further revealed that in monosyllabic conditions, no mean reaction time differences were significant between the native Mandarin and advanced L2 groups, between the native Mandarin and beginning L2 groups, and between the advanced and beginning L2 groups. In disyllabic conditions, no subject group differences reached statistical significance.

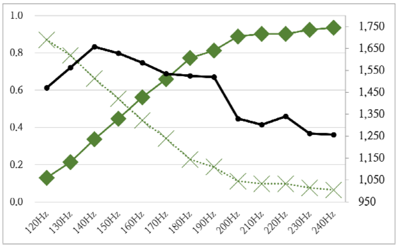
Reaction time measurements confirm the different tone identification patterns found in monosyllabic and disyllabic conditions; the stimuli were dominantly identified as Tone 1 in the former, whereas Tone 1 and Tone 3 were categorically distinguished in the latter. Figure 7 shows that reaction time decreased as pitch became higher in monosyllabic conditions, which indicates that higher pitch further facilitated Tone 1 responses. By contrast, reaction time notably increased over the identification boundary in disyllabic conditions, as presented in Figure 8. The increase was more prominent in the native Mandarin and advanced L2 groups. In Figures 7 and 8, the x-axis corresponds to pitch height, and the left and right y-axes correspond to Tone 1 response and mean reaction time (ms), respectively.



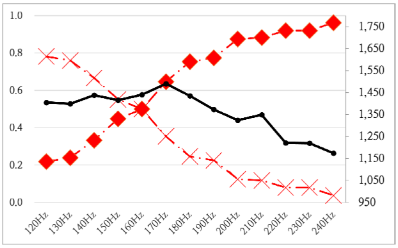
(a) native Mandarin listeners



(b) advanced L2 Mandarin listeners

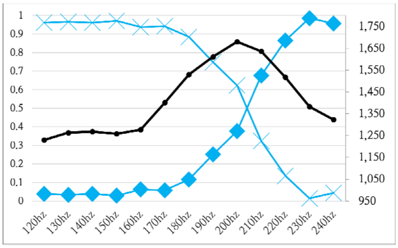


(c) intermediate L2 Mandarin listeners

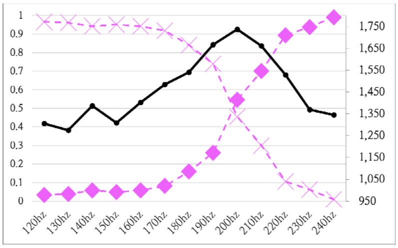


(d) beginning L2 Mandarin listeners

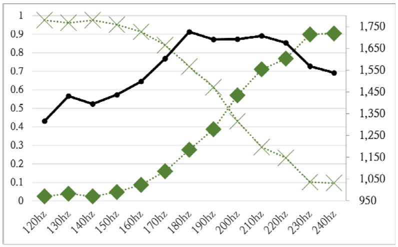
Figure 7. Mandarin tone identification and reaction time in monosyllabic conditions



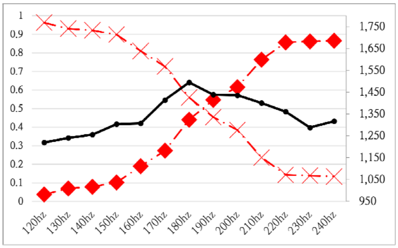
(a) native Mandarin listeners



(b) advanced L2 Mandarin listeners



(c) intermediate L2 Mandarin listeners



(d) beginning L2 Mandarin listeners

Figure 8. Mandarin tone identification and reaction time in disyllabic conditions

4.2 Experiment 2: Korean stop identification

4.2.1 Identification

Korean stop identification in monosyllabic conditions

The native Korean listeners' Korean stop identification patterns in monosyllabic conditions are presented in Figure 9. In Figures 9 to 10, the x-axis and y-axis correspond to pitch height and stop response, respectively. As fortis response was coded as '1' and lenis response was coded as '0', higher mean plots on the y-axis indicate more fortis responses. The panels from left to right represent the advanced, intermediate, and beginning L2 Mandarin groups. Recall that although Mandarin unaspirated stops are similar to Korean fortis stops in VOT, native Korean listeners who have the experience of learning Mandarin tend to associate lower pitch with lenis stops and higher pitch with fortis stops (Ko 2000; Maeng and Kwon 2008; Lee-Kim 2020). As shown in Figure 9, all three groups generally yielded more fortis responses, while they identified more heard stimuli as fortis as pitch became higher. It is noted that the advanced L2 group judged the stimuli as fortis even over a lower pitch span.

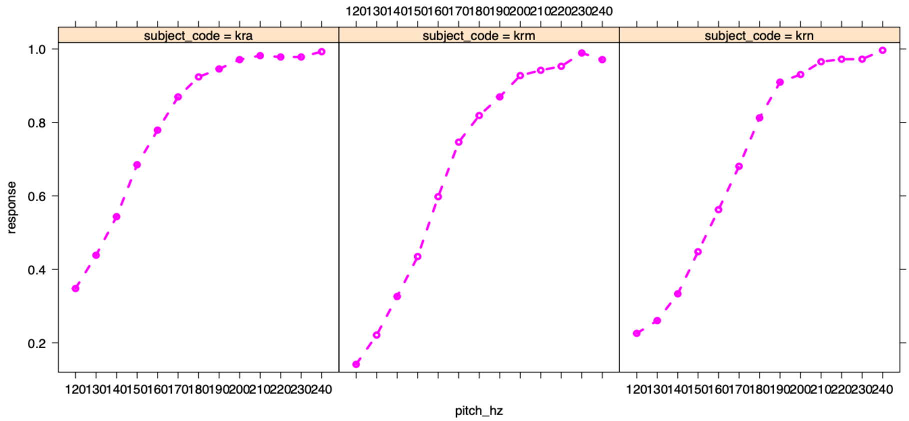


Figure 9. Korean stop identification in monosyllabic conditions
(left panel: advanced L2 Mandarin, middle panel: intermediate L2 Mandarin, right panel: beginning L2 Mandarin)

Mixed effect logistic regression model analyses found significant main effects of subject group (F=4.462, p<0.05), pitch height (F=630.585, p<0.001), onset type (F=9.2386, p<0.001), and rime type (F=28.6118, p<0.001). There were also significant interaction effects of subject group*pitch height (F=8.2362, p<0.001), subject group*onset type (F=4.0406, p<0.01), subject group*rime type (F=6.2813, p<0.001), pitch height*rime type (F=1.7166, p<0.01), onset type*rime type (F=22.8366, p<0.001), and subject group*onset type*rime type (F=2.3761, p<0.01). *Bonferroni* post hoc tests found no statistical group differences between the intermediate and beginning L2 groups. However, there were significant differences between the advanced and intermediate L2 groups over 120Hz~150Hz as well as between the advanced and beginning L2 groups over 140Hz~170Hz. This confirms that the advanced L2 Mandarin group made more fortis judgments over the lower pitch span.

Table 5 presents the percentage of fortis responses at each pitch height in the three subject groups. In Tables 5 and 6, the shaded areas indicate that the percentage changes were different at a significance level of 0.05. It shows that while the stimuli were perceived as fortis stops over a higher pitch range in all three groups, the fortis responses were distributed over a considerably wider pitch span in the advanced L2 group.

Table 5. Percentage of fortis responses in monosyllabic conditions

Pitch (Hz)	120	130	140	150	160	170	180	190	200	210	220	230	240
Kor (Adv)	34.8%	43.8%	54.3%	68.5%	77.9%	87.0%	92.4%	94.6%	97.1%	98.2%	97.8%	97.8%	99.3%
Kor (Int)	14.1%	22.1%	32.6%	43.5%	59.8%	74.6%	81.9%	87.0%	92.8%	94.2%	95.3%	98.9%	97.1%
Kor (Beg)	22.6%	26.0%	33.3%	44.8%	56.2%	68.1%	81.2%	91.0%	93.1%	96.5%	97.2%	97.2%	99.7%

Korean stop identification in disyllabic conditions

The Korean stop identification characteristics presented in Figure 10 demonstrate three major differences in lenis-fortis perception between monosyllable and disyllabic conditions (i.e., the initial syllables of disyllabic expressions). First, in contrast to the relatively gradual change from lenis to fortis responses in monosyllabic conditions, the identification shift was rather abrupt in disyllabic conditions. A quick change in lenis-fortis identification was particularly noticeable in both the advanced and intermediate L2 groups. Second, the pitch span over which lenis responses were given was considerably wider in all three groups, and the identification boundary, therefore, was moved to a higher pitch. Third, the beginning L2 group made a more gradual lenis-fortis distinction along the pitch change, albeit with fewer lenis responses over the lower pitch span. This indicates that the lenis-fortis identification was performed in a categorical manner in the advanced and intermediate L2 groups, whereas fortis responses gradually increased as pitch became higher in the beginning L2 group.

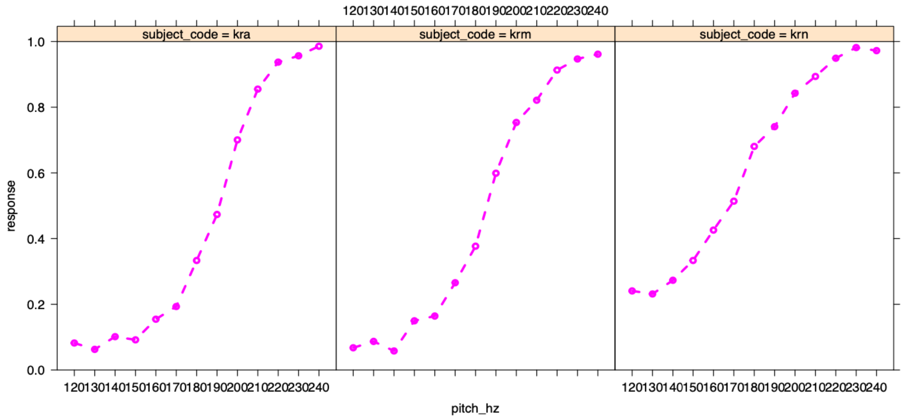


Figure 10. Korean stop identification in disyllabic conditions
(left panel: advanced L2 Mandarin, middle panel: intermediate L2 Mandarin, right panel: beginning L2 Mandarin)

Mixed effect logistic regression model analyses revealed significant main effects of subject group (F=8.203, p<0.001), pitch height (F=687.6516, p<0.001), onset type (F=11.6956, p<0.001), and rime type (F=12.6274, p<0.001). Interaction effects of subject group*pitch height (F=8.5097, p<0.001), subject group*rime type (F=2.552, p<0.001), subject group*rime type (F=6.2813, p<0.001), pitch height*rime type (F=2.552, p<0.05), onset type*rime type (F=23.2592, p<0.001), subject group*onset type*rime type(F=3.8829, p<0.001), and pitch height*onset type*rime type (F=1.697, p<0.01) were also significant. *Bonferroni* post hoc tests found no significant differences between the advanced and intermediate L2 groups. However, differences between the advanced and beginning L2 groups were significant over 150Hz~190Hz, and differences between the intermediate and beginning L2 groups were also significant over 140Hz and 160~180Hz.

Table 6 presents the percentage of fortis responses at each pitch height in the three subject groups. It confirms that the lenis-fortis identification gradually changed along with the pitch change in the beginning L2 group, while the advanced and intermediate L2 groups judged stops in a categorical fashion.

Table 6. Percentage of fortis responses in disyllabic conditions

Pitch (Hz)	120	130	140	150	160	170	180	190	200	210	220	230	240
Kor (Adv)	8.2%	6.3%	10.1%	9.2%	15.5%	19.3%	33.3%	47.3%	70.1%	85.5%	93.7%	95.7%	98.6%
Kor (Int)	6.8%	8.7%	5.8%	15.0%	16.4%	26.6%	37.7%	59.9%	75.4%	82.1%	91.3%	94.7%	96.1%
Kor (Beg)	24.1%	23.2%	27.3%	33.3%	42.6%	51.4%	68.1%	74.1%	84.3%	89.4%	94.9%	98.2%	97.2%

Korean stop identification and Mandarin tone identification

Given that the lenis-fortis identification of native Korean listeners of L2 Mandarin was influenced by the syllable pitch, an examination of the perceptual relation between the Mandarin tones and the Korean lenis-fortis contrast is in order. Figure 11 presents the overlaid patterns of the tone and stop identifications in monosyllabic conditions. In Figures 11-12, the x-axis corresponds to pitch height, and the y-axis represents Tone 1 response and fortis response. Tone 1 and fortis responses were coded as '1', and Tone 3 and lenis responses were coded as '0'. Therefore, higher mean plots drawn in the solid line and dotted line on the y-axis indicate more Tone 1 responses and fortis responses, respectively. The panels from left to right represent the advanced, intermediate, and beginning L2 groups. As presented in Figure 11, a significant mapping between the Tone 1-Tone 3 identification and the Korean lenis-fortis identification was observed over the entire pitch span in all three groups. In *Bonferroni* post hoc tests, the only significance was the difference for the pitch of 190Hz in the beginning L2 group.

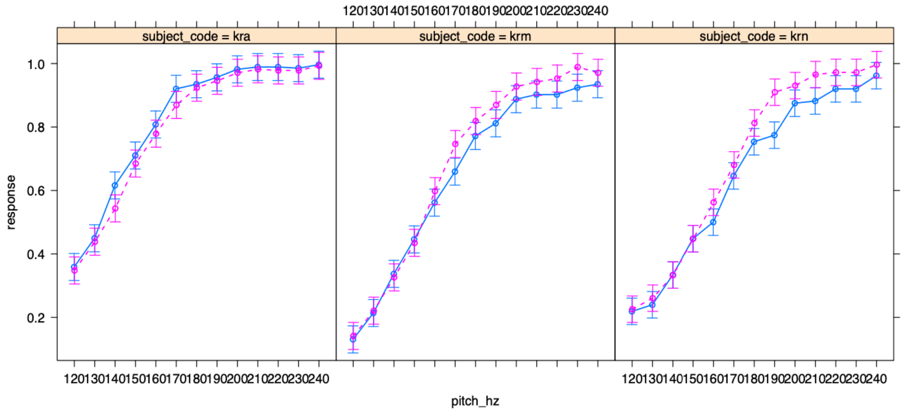


Figure 11. Korean stop identification and Mandarin tone identification in monosyllabic conditions (left panel: advanced L2 Mandarin, middle panel: intermediate L2 Mandarin, right panel: beginning L2 Mandarin; ——— tone identification, - - - - stop identification)

The overlaid patterns of the tone and stop identifications in disyllabic conditions in Figure 12 exhibit that while there were more fortis responses than Tone 1 responses at all pitch heights in the three groups, similar categorical perception characteristics were notable in the advanced and intermediate L2 groups. By contrast, the perceptual distance was readily observed in the beginning L2 group, which may be attributed to the fortis responses increasing with a higher pitch.

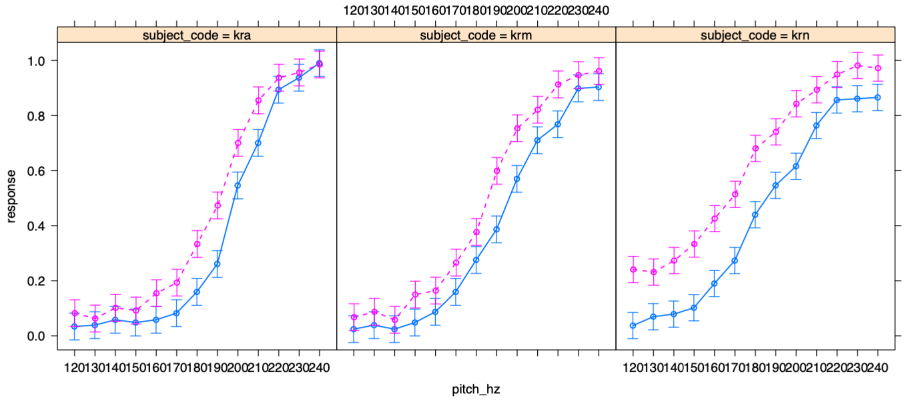
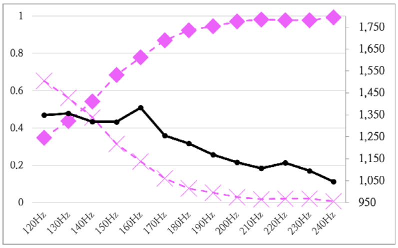


Figure 12. Korean stop identification and Mandarin tone identification in disyllabic conditions (left panel: advanced L2 Mandarin, middle panel: intermediate L2 Mandarin, right panel: beginning L2 Mandarin; —— tone identification, - - - - stop identification)

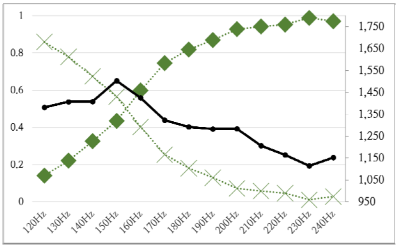
4.2.2 Reaction time

The mean reaction times for the Korean stop identification in monosyllabic conditions were '1,145 ms (beginning L2) < 1,222 ms (advanced L2) < 1,303 ms (intermediate L2)', and those in disyllabic conditions were '1,082 ms (beginning L2) < 1,341 ms (intermediate L2) < 1,354 ms (advanced L2)'. In monosyllabic conditions, reaction time differences among subject groups did not reach significance, while experiment type (i.e., tone identification vs. stop identification) had significant effects (F=205.0539, p<0.001). Interaction effects of subject group*experiment type (F=23.1082, p<0.001), experiment type*rime type (F=3.6194, p<0.05), subject group*experiment type*onset type (F=2.6758, p<0.05), subject group*experiment type*onset type*rime type (F=2.3368, p<0.01) had significant effects. In disyllabic conditions, subject group (F=245.6249, p<0.001), experiment type (F=360.5017, p<0.001), pitch (F=24.0657, p<0.001) had significant main effects, and interaction effects of subject group*experiment type (F=27.5388, p<0.001), subject group*pitch height (F=2.8098, p<0.001), experiment type*pitch height (F=6.4573, p<0.001) were also significant. *Bonferroni* post-hoc analysis results further showed that in monosyllabic conditions, no mean reaction time differences were significant between the advanced and intermediate groups, between the advanced and beginning L2 groups, and that only differences between the intermediate and beginning L2 groups were significant at the pitch of 120Hz and over 140Hz~160Hz. In disyllabic conditions, no subject group differences between the advanced and intermediate L2 groups reached statistical significance, whereas the beginning L2 group was significantly faster than the other two groups at most pitch heights.

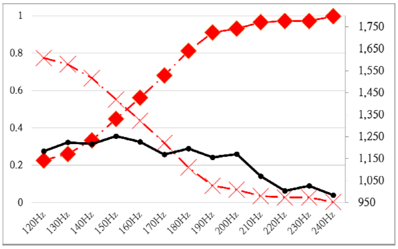
Reaction time measurements certainly demonstrate the different identification patterns observed in monosyllabic and disyllabic conditions. In both the advanced and intermediate L2 groups, the stimuli were more likely to be identified as fortis in monosyllabic conditions. And reaction time decreased as pitch became higher since fortis responses were facilitated by higher pitch. In disyllabic conditions, by contrast, the categorical judgments between lenis and fortis were made, and reaction time notably increased over the perceptual boundary. The reaction time increase was the most prominent in the advanced L2 group, whereas the categorical identification with increasing reaction time over the categorical boundary was realized to the least extent in the beginning L2 group. Figures 13 and 14 present the relations between the lenis-fortis identification and reaction time in monosyllabic conditions and those in disyllabic conditions, respectively. In the figures, the x-axis corresponds to pitch height, and the left and right y-axes correspond to fortis response and mean reaction time (ms), respectively.



(a) advanced L2 Mandarin group

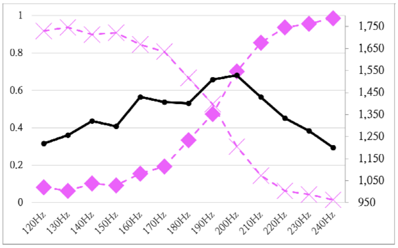


(b) intermediate L2 Mandarin group

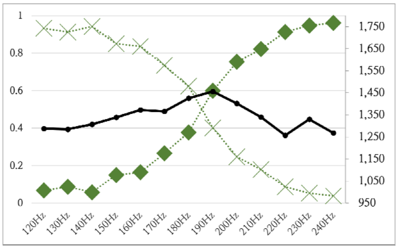


(c) beginning L2 Mandarin group

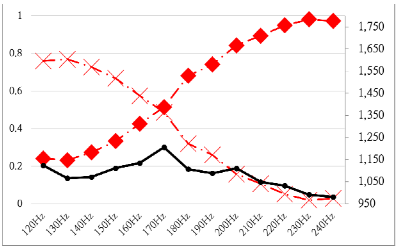
Figure 13. Korean stop identification and reaction time in monosyllabic conditions



(a) advanced L2 Mandarin group



(b) intermediate L2 Mandarin group



(c) beginning L2 Mandarin group

Figure 14. Korean stop identification and reaction time in disyllabic conditions

5. Discussion

Results of the experiments conducted for the present study show that the Mandarin level tones (i.e., Tone 1 and Tone 3) can be categorically identified by both native and non-native listeners. The perceptual patterns, however, interact with the condition in which tone occurs as well as with non-native listeners' L2 proficiency. In monosyllabic conditions, native Mandarin listeners exhibit a strong tendency to identify the tones of differing pitch heights as Tone 1. This can be attributed to the phonetic constraint that a low level tone does not occur in isolation but precedes other tones (i.e., Tones 1, 2, and 4), whereas a high level tone has no such constraint. Therefore, the level tone in monosyllabic conditions is likely to be judged as Tone 1 with little regard to pitch heights. By contrast, it is apparent that native Mandarin listeners identify Tone 1 and Tone 3 in a categorical manner on the initial-syllable in disyllabic conditions. It should be noted that while the syllable can be either Tone 1 or Tone 3 in the Mandarin phonology, more Tone 3 judgments are made. This finding is in line with earlier studies, some of which performed an experiment that required listeners to choose between real words of differing tones (Fox and Qi 1990; Moore and Jongman 1997; Cao 2010 a, b; Sjerps et al. 2018; Shi 2019). Considering that the second syllables in disyllabic conditions were designed to be Tone 1, the considerable increase of Tone 3 judgments should be ascribed to the contrast effect of the anchoring high tone on the following syllable.

A further examination of the present study finds that there is a great deal of perceptual similarity between native Mandarin listeners and native Korean listeners of advanced L2 Mandarin. In both monosyllabic and disyllabic conditions, no significant differences are found between the two groups of listeners. However, less proficient L2 Mandarin listeners' tone judgments appear to be generally shaped by acoustic pitch property in monosyllabic conditions; intermediate and beginning L2 listeners' Tone 1 identification increases as pitch becomes higher. In disyllabic conditions, non-native listeners also show a categorical tone identification although the boundary slopes are less sharply defined in intermediate and beginning listeners than in advanced L2 listeners. No significant differences are found between the intermediate and beginning L2 listener groups. While the contrast effect of the anchoring tone is significant in both native and non-native listeners' tone perception, no effects of onset consonant or rime type are found. Figure 15 presents the schematized comparisons of the native and non-native tone perception patterns observed in the present study.

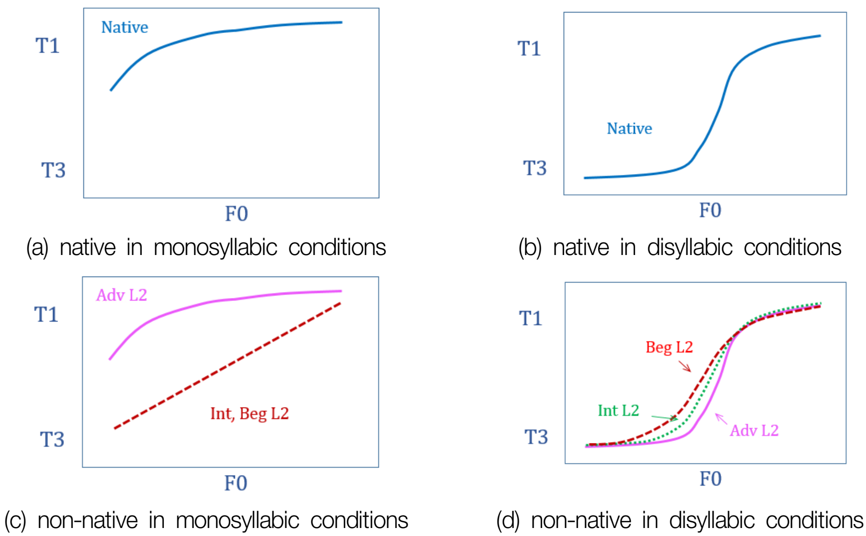
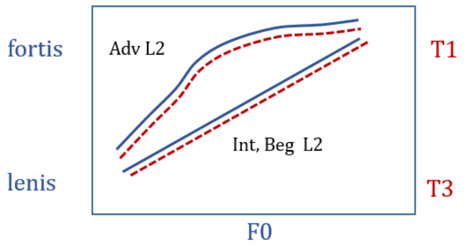


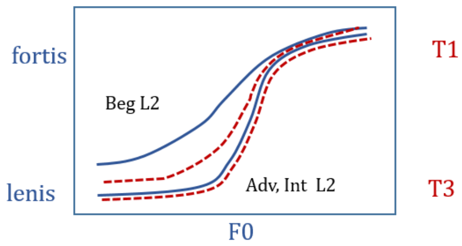
Figure 15. Native and non-native tone identification comparisons

Overall, reaction time measurements support a non-categorical identification in monosyllabic conditions and a categorical identification in disyllabic conditions in both native and non-native listeners. That is, reaction time decreases as pitch becomes higher in monosyllabic conditions and notably increases over the categorical boundary in disyllabic conditions. Interestingly, native listeners are not necessarily faster than non-native listeners in tone identification. Furthermore, no correlation is found between reaction time and L2 proficiency. In general, the shortest reaction time is found in beginning L2 listeners.

Results of the native Korean listeners' Korean lenis-fortis stop identification further reveal an intriguing relation between L1 segmental identification and L2 tone identification. In monosyllabic conditions, while more fortis judgments are made in general, notable perceptual mappings between the two types of identification are found in native Korean listeners. In disyllabic conditions, both the advanced and intermediate L2 groups show a categorical identification with a steeper boundary slope realized in the former. By contrast, the perceptual patterns are matched to the least extent in the beginning L2 group. This indicates that despite the same VOT, the native Korean listeners' lenis-fortis identification is considerably influenced by the syllable pitch. In other words, a reorganization of perceptual cues is witnessed; VOT cues become overweighted by F0 cues in stop perception of native Korean listeners of L2 Mandarin. Figure 16 captures the stop-tone perceptual mapping patterns of the native Korean listener groups found in this study. In the figure, the solid line and dashed line represent the lenis-fortis identification and Tone 1-Tone 3 identification, respectively.



(a) monosyllabic conditions



(b) disyllabic conditions

Figure 16. Stop and tone identification mapping comparisons

Reaction time measurements confirm a categorical lenis-fortis identification in disyllabic conditions. The most sharply defined categorical boundary in the advanced L2 group is also supported by the larger reaction time realized over the lenis-fortis boundary, while the increase of reaction time over the perceptual boundary in the beginning L2 group is considerably smaller than those in the advanced and intermediate L2 groups. This is in contrast to the reaction times decreasing along with a higher pitch in monosyllabic conditions in all three L2 groups. In general, the shortest reaction time is found in the beginning L2 group, which may indicate that their sound perception is likely shaped by acoustic correlates with less phonological interference between L1 and L2 due to relatively limited L2 experience.

6. Conclusion

The present study investigated how native and non-native listeners of Mandarin perceive the Mandarin level tones (i.e., Tone 1 and Tone 3) and how the Mandarin tone identification interacts with the Korean lenis-fortis stop perception in native Korean listeners of L2 Mandarin. Two perception experiments, namely *Mandarin Tone 1-Tone 3 Identification Experiment* and *Korean Lenis-Fortis Identification Experiment*, were conducted. Results show that regardless of pitch heights, native Mandarin listeners tend to identify the monosyllabic level tones as Tone 1, whereas the level tones on the first syllable in disyllabic expressions are categorically identified as either Tone 1 and Tone 3. This shows that, just as in segmental sounds, a quantal relation between articulatory and acoustic parameters is found in tones of varying F0 cues (Liberman 1970; Stevens 1989, 1998). The perceptual difference between monosyllabic conditions and disyllabic conditions should be attributed to the phonology of Mandarin and the perceptual effect of the following tone. Due to the phonological constraint that does not allow Tone 3 to surface as low tone in isolation in Mandarin, a tone of level pitch that occurs in monosyllabic condition generally corresponds to Tone 1 with little regard to pitch heights. With no such constraint in disyllabic conditions, the level tones preceding Tone 1 can be heard as either Tone 1 or Tone 3. And the tendency to identify the tone as Tone 3 more frequently in disyllabic conditions should be explained by the contrast effect of the anchoring high tone. Listeners tend to hear tones as low when the adjacent tone is high, while tones are identified as high when the neighboring tone is low (Lin and Wang 1984; Moore and Jongman 1992). It is further demonstrated that the proficiency of L2 Mandarin listeners exerts a strong influence on tone perception; the advanced L2 listeners' responses are generally in line with those of the native Mandarin listeners in both monosyllabic and disyllabic conditions. While less proficient listeners' tone judgments in monosyllabic conditions considerably correlate with the pitch height change, a tendency to categorically distinguish Tone 1 and Tone 3 emerges in disyllabic conditions yet with less well defined perceptual boundaries in beginning L2 listeners.

Another important finding of the present study pertains to the L2 tone effect on L1 consonant perception. Despite the fact that Mandarin unaspirated stops resemble Korean fortis in VOT, F0 cues are utilized for matching Mandarin stops to Korean stops by native Korean listeners of L2 Mandarin. More importantly, native Korean listeners distinguish lenis and fortis in a categorical manner in disyllabic conditions when only F0 cues and no VOTs are changed, and a large similarity is found between Korean fortis-lenis perception and Mandarin Tone 1-Tone 3 identification particularly in relatively proficient L2 listeners (i.e., advanced and intermediate L2 listeners). The perceptual mapping of L2 Mandarin unaspirated stops to two stop categories (i.e., lenis and fortis) in native Korean listeners demonstrates that L2 listeners develop distinct phonetic cues in L2 sound perception and that L2 proficiency has a significant influence on reshaping the phonological categories. The findings of the present study therefore contribute to expanding our knowledge on L2 sound perception in two important respects. First, while a handful of studies have observed an influence of pitch on Mandarin stop perception in native Korean listeners (Ko 2000; Maeng and Kwon 2008; Lee-Kim 2020), the present study further discovers that the rephonologization of L2 stop categories changes as L2 proficiency advances and that the extensive experience with L2 can replace the phonetic cues used in L1 (i.e., VOT) with those utilized in L2 (i.e., F0). Second, with a vast majority of L2 sound perception research focusing on the perceptual challenges with segmental sounds that either do not exist in L1 or do not make a one-to-one phonological mapping between L1 and L2, this study presents a new finding that the knowledge of phonetic properties and phonological constraints of L2 prosody has a significant influence on the rephonologization of segmental sounds.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects

models using lme4. *Journal of Statistical Software* 67(1): 1-48.
Belotel-Grenié, Agnès and Michel Grenié. 1994. Phonation types analysis in standard Chinese. *Proceedings of 3rd International Conference on Spoken Language Processing*, 343-346. International Speech Communication Association.
Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171-206. Baltimore: York Press.
Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America* 109(2): 775-794.
Best, Catherine T. and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn and Nurray J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 13-34. Amsterdam: John Benjamins.
Cao, Jianfen and Maddieson Ian. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* 20(1): 77-92.
Cao, Wen. 2010a. A study on Chinese level tone perception [in Chinese]. *Zhongguo Yuwen* 339: 536-576.
Cao, Wen. 2010b. A comparative study on tone perception – Report on level pitch contours [in Chinese]. *Shijie Hanyu Jiaoxue* 24: 255-262.
Chao, Yuen Ren. 1930. A system of tone letters. *Le Maître Phonétique* 45: 24-27.
Chao, Yuen Ren. 1932. A preliminary study of English intonation (with American variants) and its Chinese equivalents. *Ts'ai Yuan P'ei Anniversary Volume (Supplementary volume 1 of the Bulletin of the Institute of History and Philology)*, 105-156. Taipei: Institute of History and Philology, Academia Sinica.
Chao, Yuen Ren. 1933. A preliminary study of English intonation (with American variants) and its Chinese equivalents. *Bulletin of the Institute of History and Philology* 4: 121-134.
Chao, Yuen Ren. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In Morris Hallé (ed.), *For Roman Jakobson: Essays on the occasion of his sixtieth birthday*, 52-59. The Hague: Mouton.
Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
Chao, Yuen Ren, Changpei Luo, and Fanggui Li. 1948. *Studies of Chinese phonology.* Shanghai: Commercial Press.
Chen, Chung-yu. 1984. Neutral tone in Mandarin: Phonotactic description and the issue of the norm. *Journal of Chinese Linguistics* 12(2): 299-333.
Cheng, Chin-Chuan. 1973. *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.
Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30(2): 193-228.
Duanmu, San. 2007. *The phonology of standard Chinese* (Second Edition). Oxford: Oxford

University Press.

Fox, Robert A. and Ying-Yong Qi. 1990. Context effects in the perception of lexical tone. *Journal of Chinese Linguistics* 18(2): 261-284.

Fu, Yuping and Yong-cheol Lee. 2022. The production of tone 3 by advanced Korean learners of Mandarin. *Linguistic Research* 39(1): 213-233.

Gao, Jianyin and Pierre Hallé. 2015. The role of voice quality in Shanghai tone perception. *Proceedings of the 18th International Congress of Phonetic Sciences*, 1-5. International Congress of Phonetic Sciences.

Gao, Jianyin and Pierre Hallé. 2017. Phonetic and phonological properties of tones in Shanghai Chinese. *East Asian Languages and Linguistics* 46(1): 1-31.

Han, Jeong-Im. 1996. *The phonetics and phonology of "tense" and "plain" consonants in Korean*. PhD Dissertation. Cornell University.

Huang, Tsan and Keith Johnson. 2010. Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica* 67(4): 243-267.

Huang, Jingyuan and Lori L. Holt. 2009. General perceptual contributions to lexical tone normalization. *The Journal of the Acoustical Society of America* 125(6): 3983-3994.

Jang, Hyejin. 2012. *Acoustic properties and perceptual cues of Korean word-initial obstruents*. PhD Dissertation. Korea University.

Jeong, Yun Ja. 2010. Speech perception of Korean plain, aspirated, and tense considering pitch of the following vowel [in Korean]. *Urimal* 27: 73-94.

Joh, Geol. 1996. Jung-Han songgieum daebi sogo [in Korean]. *Proceedings of Seoul International Conference on Phonetic Sciences*, 277-284. Korean Society of Speech Sciences.

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.

Kim, Hyunsoon, Shinji Maeda, Kiyoshi Honda, and Lise Crevier-Buchman. 2017. The mechanism and representation of Korean three-way phonation contrast: external photoglottography, intra-oral air pressure, airflow, and acoustic data. *Phonetica* 75(1): 57-84.

Kim, Midam. 2004. Correlation between VOT and f0 in the perception of Korean stops and affricates. *Proceedings of the 8th International Conference on Spoken Language Processing*, 49-52. International Speech Communication Association.

Kim, Mi-Ryoung. 2000. *Segmental and tonal interactions in English and Korean: A phonetic and phonological study*. PhD Dissertation. University of Michigan, Ann Arbor.

Kim, Mi-Ryoung. 2014. Ongoing sound change in the stop system of Korean: A three- to two-way categorization. *Studies in Phonetics, Phonology, and Morphology* 20(1): 51-82.

Ko, Eon-Suk. 2003. The laryngeal effect in Korean: Phonology or phonetics?. In Jeroen van de Weijer, Vincent J. van Heuven, and Harry van der Hulst (eds.), *The phonological spectrum volume 1: Segmental structure*, 171-191. Amsterdam: John Benjamins.

Ko, Mi-Sook. 2000. A study on the relationship between segments and suprasegmental features [in Korean]. *Korea Journal of Chinese Linguistics* 12: 235-256.

Kuang, Jianjing. 2013a. *Phonation in tonal contrasts*. PhD Dissertation. University of California, Los Angeles.

Kuang, Jianjing. 2013b. The tonal space of contrastive five level tones. *Phonetica* 70(1-2): 1-23.

Lee, Chao-Yang, Liang Tao, and Zinny S. Bond. 2009. Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics* 37(1): 1-15.

Lee, Sun Woo and Tae Eun Kim. 2022. A study of the influence of native phonological system on the phonetic perception and production [in Korean]. *The Journal of Chinese Language and Literature* 134: 7-28.

Lee, Yuh-Shiow, Douglas A. Vakoch, and Lee H. Wurm. 1996. Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research* 25(5): 527-542.

Lee-Kim, Sang-Im. 2020. Tone-induced split in stop category mapping by Korean learners of Mandarin Chinese. *Language and Linguistics* 21(1): 54-79.

Lenth, Russell V. 2022. emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.2. https://CRAN.R-project.org/package=emmeans

Liberman, Alvin M. 1970. Some characteristics of perception in the speech mode. *Association For Research in Nervous and Mental Disease* 48: 238-254.

Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.

Lin, Tao and William S.-Y. Wang. 1984. The issue of tone perception [in Chinese]. *Zhongguo Yuyan Xuebao* 2: 59-69.

Liu, Hanjun, Manwa L. Ng, Mingxi Wan, Supin Wang, and Yi Zhang. 2008. The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice* 22(2): 210-218.

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

Maddieson, Ian. 1978. Universals of tone. In Joseph Greenberg (ed.), *Universals of human language vol.2: Phonology*, 335-363. Stanford: Stanford University Press.

Maeng, Joo-Oeck and Yongsil Kwon. 2007a. A Contrastive acoustic analysis on the unvoiced velar of Korean and Chinese [in Korean]. *The Journal of Chinese Language and Literature* 25: 499-523.

Maeng, Joo-Oeck and Yongsil Kwon. 2007b. A contrastive acoustic analysis on the unvoiced alveolar plosive of Korean and Chinese [in Korean]. *The Journal of Chinese Studies* 41: 81-204.

Maeng, Joo-Oeck and Yongsil Kwon. 2008. A perception test of the unvoiced plosive of Chinese before and after changing fundamental frequency [in Korean]. *The Journal of Chinese Studies* 44: 73-92.

Moore, Corrine B. and Allard Jongman. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America* 102(3): 1864-1877.

Peng, Gang, Caicai Zhang, Hong-Ying Zheng, James W. Minett, and William S.-Y. Wang. 2012. The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research* 55(2): 579-595.

Peng, Jui-Feng, Li-mei Chen, and Chia-Cheng Lee. 2009. Tonal effects on voice onset time. *International Journal of Computational Linguistics and Chinese Language Processing* 14(4): 341-361.

Peng, Shu-hui. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. In Michael B. Broe and Janet B. Pierrehumbert (eds.), *Acquisition and the lexicon: Papers in laboratory phonology V*, 152-167. Cambridge: Cambridge University Press.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/

Ren, Nianqi. 1992. *Phonation types and stop consonant distinctions: Shanghai Chinese*. PhD Dissertation. University of Connecticut, Storrs.

Schertz, Jessamyn, Taehong Cho, Andrew Lotto, and Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52: 183-204.

Shi, Feng. 2019. *Auditory patterns: A preliminary study on perception features of Chinese speech perception* [in Chinese]. Beijing: The Commerce Press.

Shi, Feng, and Qibin Ran. 2011. A comment on "Mandarin tone perception: A report on a low level tone" [in Chinese]. *Studies of the Chinese Language* 6: 550-555.

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

Sjerps, Matthias J., Caicai Zhang, and Gang Peng. 2018. Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context. *Journal of Experimental Psychology: Human Perception and Performance* 44(6): 914-924.

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

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17(1): 3-45.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge: The MIT Press.

Wang, Li. 1989. Some problems in phonetic analysis of Modern Chinese [in Chinese]. In Li Wang (ed.), *Collected works of Wang Li volume 17*, 32-45. Shandong: Shandong Education Express.

Wang, William S.-Y. and Kung-pu Li. 1967. Tone 3 in Pekinese. *Journal of Speech and Hearing Research* 10(3): 629-636.

Yang, Ruoxiao. 2011. The phonation factor in the categorical perception of Mandarin tones. In *Proceedings of 17th International Congress of Phonetic Sciences*, 2204-2207. International Congress of Phonetic Sciences.

Yang, Ruoxiao. 2015. The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics* 43(1): 453-472.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yuan, Jiahong and Yiyan Chen. 2014. 3rd tone sandhi in Standard Chinese: A corpus approach. *Journal of Chinese Linguistics* 42(1): 218-237.

Zee, Eric. 1980a. A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Phonetics* 49: 98-116.

Zhang, Caicai, Gang Peng, and William S.-Y. Wang. 2012. Unequal effects of speech and nonspeech contexts on perceptual normalization of Cantonese level tones. *The Journal of the Acoustical Society of America* 132(2): 1088-1099.

Zhang, Jie. 2010. Issues in the analysis of Chinese tone. *Language and Linguistics Compass* 4(12): 1137-1153.

Zhang, Yanan and Chin-shik Kim. 2019. A study on the Korean and Chinese plosive sounds comparison based on acoustic and auditory experiments [in Korean]. *Studies in Linguistics* 52: 155-171.

Kyungmin Lee
PhD Candidate
Department of Chinese Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826 Korea
E-mail: leegmf4@snu.ac.kr

Ok Joo Lee
Associate Professor
Department of Chinese Language and Literature
Seoul National University
1 Gwanak-ro, Gwanak-gu,
Seoul, 08826 Korea
E-mail: ojlee@snu.ac.kr

Received: 2022. 10. 04.
Revised: 2022. 11. 15.
Accepted: 2022. 11. 18.