

EFFECTS OF ORTHOGRAPHIC INPUT ON L2 PRODUCTION: THE CASE OF KOREAN-SPEAKING LEARNERS OF MANDARIN CHINESE

National Chiao Tung University
Sang-Im Lee-Kim
sangimleekim@nctu.edu.tw

ABSTRACT

This work presents a longitudinal study of the role of orthographic input in the development of Mandarin stops and lexical tones by Korean speakers. One group of learners was tested using Pinyin, a Romanized representation of Mandarin, while the other group was tested with auditory stimuli. It was shown that Pinyin-mediated Mandarin voiceless unaspirated stops (e.g., <b d>) may be incorrectly produced with long aspiration followed by a low-*f0* tone due to its mapping onto Korean lenis stops [Long VOT–Low *f0*]. Further, more L2 experience may lead to poorer production of the T2 (<bá>) - T3 (<bǎ>) contrast, which is indicative of the pervasive role of the knowledge of the L2 orthography and sound system during early L2 acquisition. The results of the study contribute empirical evidence for the interference of L2 orthography at the subphonemic level to the growing body of research on orthographic input and L2 phonological acquisition.

Keywords: Tone, stops, L1-L2 category mapping, orthography, Mandarin

1. INTRODUCTION

Of the many factors affecting the acquisition of L2 phonology, the role of orthographic input has only recently been considered. Despite some empirical findings demonstrating a positive role of orthographic input in L2 phonological acquisition [10, 12], other experimental results have suggested a negative influence [5, 13]. In particular, when orthographic input is not entirely consistent with the corresponding phonetic forms, phonological development may be hindered. For example, in phoneme counting and segmentation tasks, native English speakers learning Mandarin based on Pinyin, a Romanized orthographic representation, could correctly count the number of segments in a syllable when the sounds matched the Pinyin (e.g., [iou] <you>), but when the main vowel was not represented in the Pinyin (e.g., [li(o)u], <liu>), they missed the unrepresented segment [1].

Beyond the segmental domain, significant influences of orthographic input on tone acquisition have been reported as well. In a novel word-learning study [10], naïve English speakers learning Mandarin with Pinyin tone marks (<mā má mǎ mà>)

outperformed a group of naïve speakers learning Mandarin with auditory input only. Interestingly, however, a negative impact of Pinyin input was found in a study with experienced learners of Mandarin. Cantonese-speaking advanced learners of Mandarin performed poorly when presented with Pinyin tone marks compared to when presented with logographic Chinese characters lacking tone marks, particularly for multisyllabic words [6]. The seemingly contradictory results of the two studies are suggestive of possible dynamic changes in the role of orthographic input over the course of L2 acquisition. In early stages of learning, when distinct tone categories need to be established, orthographic input may be beneficial; however, the symbolic tone representations may become rather misleading when learners need to master more fine-grained phonetic features of tonal targets.

While most studies investigating effects of orthographic input on phonological acquisition have focused on psycholinguistic and/or perceptual aspects of learning, the present study is primarily concerned with L2 production to shed light on the development of the fine-grained sub-phonemic characteristics of non-native sounds. Specifically, we explore the development of stop categories and lexical tones in Mandarin by native speakers of Seoul Korean whose native language has different sets of stops and lacks lexical tone. The learners' development was closely tracked in a longitudinal study. To address the role of orthographic input, one group of learners was tested using orthographic representations, while the other group was tested with auditory stimuli.

2. METHOD

2.1. Participants

Fourteen Seoul Korean speakers with no prior experience with tone languages participated in the production study and received monetary compensation. The participants were randomly divided into two groups based on the kind of stimuli presented in the experiment, orthographic (“Pinyin group”) or audio (“audio group”). Eight participants (5F, 3M) were assigned to the Pinyin group, six (2F, 4M) to the audio group. All participants were college students between 20 and 28 years old ($M= 25.8$). Though none had studied Chinese, all of the participants had taken compulsory English courses in primary and secondary school.

Participants were students who had registered for Elementary Chinese courses to fulfill general education requirements. The courses took place in a formal class setting four hours a week, including three hours of lectures taught in Korean by a Korean instructor and a one-hour practicum taught in Mandarin by a Chinese assistant. Although students came from different classes, all Elementary Chinese courses used the same textbook and audio files and maintained the same course structure closely designed and monitored by the Chinese department. The first week of the course was devoted to familiarizing the students with Pinyin, with a special focus on the four lexical tones (e.g., <mā má mǎ mà>, corresponding to T1(X⁵⁵) T2(X³⁵) T3(X²¹⁴) T4(X⁵¹), respectively).

2.2. Stimuli and Procedure

The stimuli were monosyllabic words consisting of Mandarin voiceless stops contrasting in aspiration followed by the vowel /a/ (e.g., unaspirated: <bā bá bǎ bà> vs. aspirated: <pā pá pǎ pà>). A total of 24 stimuli (3 (unasp, asp, nasal) x 2 (lab, cor) x 4 (tones)) were repeated three times in each task, and the same materials were used throughout the longitudinal study. Participants were tested three times over a semester: Phase 1 (after one week of Mandarin exposure), Phase 2 (after 8 weeks), and Phase 3 (after 15 weeks). Three participants (1F, 2M) from the Pinyin group and one female participant from the audio group did not participate in the last phase of the experiment.

For the Pinyin group, stimuli were presented orthographically without auditory stimuli. In each trial, a stimulus item was presented on a computer screen, and participants read the word aloud, proceeding at their own pace. Participants in the audio group were presented with auditory stimuli produced by a female native speaker of Mandarin. This experiment was run in E-prime, and participants repeated what they heard without visual stimuli. The recordings of all participants were made in a sound-attenuated booth using a Zoom H4 recorder connected to a Shure SM58 microphone at a 44 kHz sampling rate.

2.3. Acoustic analysis

The recordings were first annotated in Praat, specifically (1) the onset of the stop burst, (2) the onset of vowel voicing, and (3) the end of voicing. Stop VOT was measured by subtracting (1) from (2), and *f0* values were measured between (2) and (3) using the STRAIGHT algorithm [6] in VoiceSauce. Octave errors were corrected manually.

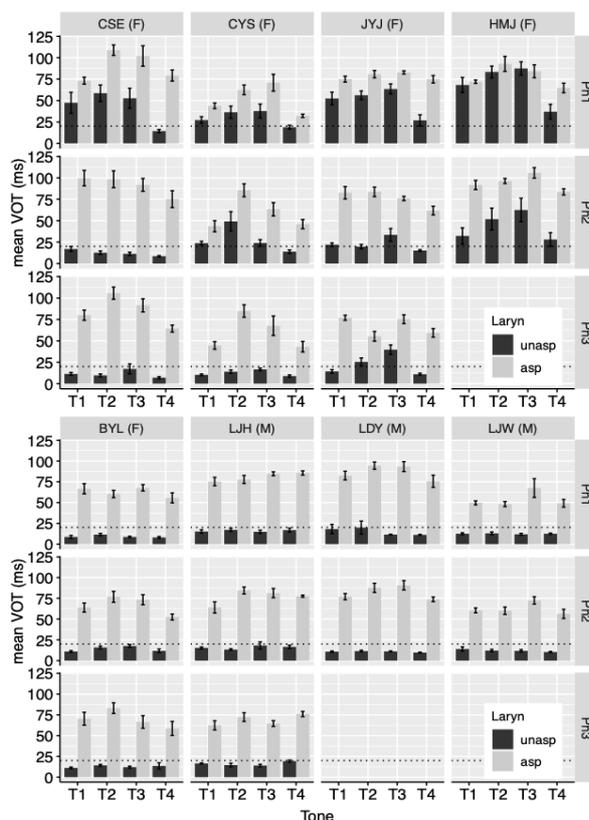
3. RESULTS

3.1. The association between stop VOT and tone

3.1.1. Stop production

All participants correctly produced long VOTs for the aspirated stops, although the values were on average shorter ($M=73$ ms) than Mandarin norms ($M=103$ ms) [8]. As for the unaspirated stops, four out of the eight speakers in the Pinyin group produced inordinately long VOTs (Figure 1, top), especially in the earlier stages of learning. While they generally made significant improvements over time, individuals progressed at different speeds. Only one participant had adjusted VOT values by Phase 2 (CSE). Some continued producing long VOTs far into the semester (CYS, HMJ), while others did not master stop targets in certain tones within the semester (JYJ). The four other learners in the Pinyin group produced correct VOT targets for the unaspirated stops from the beginning stage and maintained them throughout (Figure 1, bottom).

Figure 1: Pinyin-based learners showing inordinately long VOTs (top) vs. short VOTs (bottom) for the Mandarin unaspirated stops.



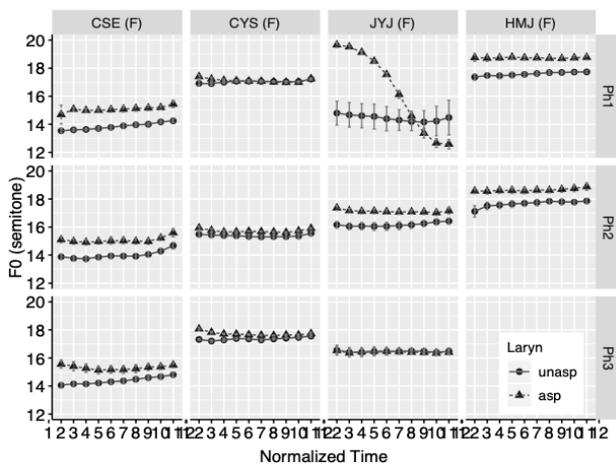
In contrast to the Pinyin group, none of the six learners in the audio group demonstrated erroneous VOT patterns; they all correctly produced long VOTs for the aspirated stops and short VOTs for the unaspirated stops. A Chi-square test comparing task conditions and the frequency of VOT overestimation

had a significant result ($\chi^2(1, N=14)=4.2, p=.040$), suggesting that the group difference can be attributed to the type of input in the tasks.

3.1.2. T1 production

T1 production was analyzed because, compared with other contour tones, the level tone was less likely to cause difficulty for the Korean learners. Overall, a clear correlation was observed between the erroneous patterns in VOT production and those in f_0 : those with overestimated VOTs for the unaspirated stops tended to produce the stops with lower f_0 throughout the entire syllable (Figure 2). Regression analyses showed that the f_0 at the vowel midpoints was significantly different for those who produced elongated VOTs (with the exception of CYS Ph1 and JYJ Ph3). In contrast, those who produced accurate stop VOTs showed no such separation in T1 f_0 trajectories between aspirated and unaspirated stops, nor did any of the participants in the audio group.

Figure 2: Mean f_0 trajectories of T1 for individual speakers with VOT overestimation for the Mandarin unaspirated stops.



Notice, however, that improvement in stop VOT production did not necessarily occur concurrently with T1 development. For some learners, f_0 differences for T1 continued even after the learning to reduce the VOT (CSE, CYS). For JYJ, however, the opposite trend was observed—VOT differences persisted into Ph3 after learning to properly produce T1 (this learner mistakenly produced T4 for aspirated stops in Ph1 and low f_0 for unaspirated stops in Ph2).

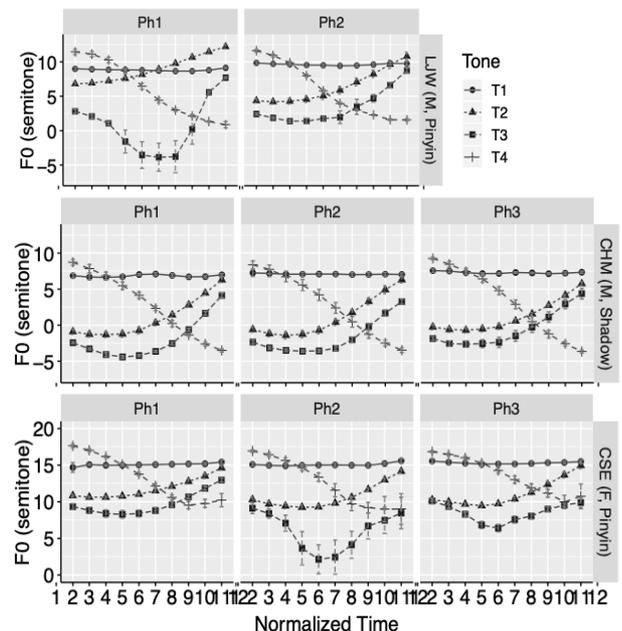
3.2. T2-T3 contrasts

The f_0 trajectories of T1 and T4 were acceptable for all participants from the very beginning stage and they continued (Figure 3). However, T2 and T3 presented a challenge, consistent with previous studies showing nonnative listeners' difficulty to perceive the difference between these two tones [2].

Although learners were aware of the two distinct categories, as evidenced by the slight but consistent divergence between the two f_0 trajectories, the production of the two tones proved problematic. Unlike native Mandarin speakers' production wherein T2 and T3 f_0 trajectories diverge to a greater extent toward the end of a syllable [4], the learners' T2 and T3 f_0 trajectories tended to be more similar.

Interestingly, many learners' ability to accurately produce T2 and T3 diminished over time, rather than improving. Figure 3 illustrates some representative cases, and Table 1 summarizes individual performances, with indications of whether or not they produced the T2-T3 contrast and the direction of their progress.

Figure 3: Mean f_0 trajectories of four lexical tones following the aspirated stops. Representative cases illustrating the (partially) diminishing T2-T3 contrasts.



Compared with the differences in VOT production between the two groups, the differences in tonal production were less clear-cut. However, more diverse patterns were found in the Pinyin group: the tonal contrast initially increased then diminished (CSE, CYS), the two tones were completely merged (BYL), and sometimes categorically wrong tonal targets were produced (LDY, JYJ). In contrast, no participant in the audio group displayed completely erroneous patterns. Without on-line auditory stimuli, learners in the Pinyin group seem more liable to implement unconstrained tonal targets in their production.

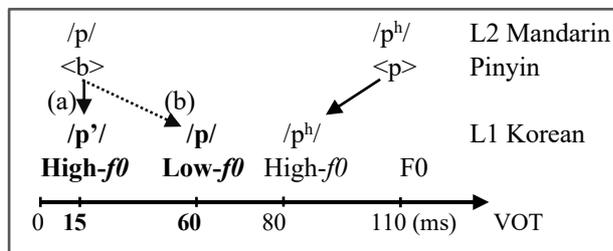
Table 1: Summary of the patterns of T2-T3 contrasts by individual learners. The arrows indicate the direction of changes in $\Delta(T2-T3)$ over time, and the gray indicates the learners showing diminishing trends.

Pinyin		Audio	
LJH	√, no change	KMJ	√, no change
HMJ	√, no change	CHM	√, ↓
CSE	√, ↑↓	LHW	√, ↓
CYS	√, ↑↓	KHB	marginal √, ↓
LJW	√, ↓	YYS	merge (Ph1), marginal √ (Ph2, 3)
LDY	marginal √ (Ph1), T3 uncategorized (Ph2), ↓	KKH	merge (Ph1), marginal √ (Ph2, 3)
JYJ	marginal √ (Ph1), T2/T1 merge (Ph2, 3)		
BYL	merge (all phases)		

4. DISCUSSION

The results of the stop production study suggest that beginner learners make a one-to-one correspondence between L2 and L1 stop categories, which are, in turn, straightforwardly reflected in production as illustrated in Figure 4. In particular, the learners who overestimated VOTs were likely to have mapped Mandarin unaspirated stops onto Korean lenis stops [Long VOT–Low f_0] (4.b), while those with shorter (i.e., more accurate) VOTs mapped them onto Korean fortis stops [Short VOT–High f_0] (4.a). This mapping is evidenced by the low- f_0 values appearing predominantly after the unaspirated stop target with elongated VOTs.

Figure 4: L1-L2 stop category mapping



Importantly, while signs of both the unaspirated-to-lenis mapping (4.b) and the unaspirated-to-fortis (4.a) mapping were observed for the learners in the Pinyin group, only the unaspirated-to-fortis mapping (4.a) was attested for the learners in the audio group. This result highlights the negative influence of orthographic input on the production of L2 sounds. Korean learners of English have been reported to map English voiced stops <b, d> onto lenis stops more so than onto fortis stops [7, 9]. As all Korean participants in the present study were familiar with the English alphabet, Pinyin orthography (e.g., <b, d>) may trigger a mapping onto lenis stops [+aspiration], overriding actual acoustic-phonetic characteristics of Mandarin unaspirated stops [-aspiration]. Without the interference of orthographic input and with the help of on-line auditory input, the learners in the audio group were not liable to such

production errors. Pedagogically, this suggests that much caution should be made when using Pinyin for learners who are familiar with English orthography. The effects of alphabetical representations could be pervasive, negatively impacting the production of both stop VOTs and tones in Mandarin.

Unlike previous studies focusing on the adverse effect of the knowledge of L2 phonology in the perceptual domain, the current study further shows that “some” L2 experience in the early phonological development may have a negative impact on production. For the difficult T2 vs. T3 contrast in Mandarin, some Korean learners managed to produce the different contours in the early stage of learning, but with more experience, the contrast may become less robust. This result is particularly interesting for the audio group. The auditory stimuli were identical in each of the three test phases, but some learners’ performance still diminished even when full auditory information was provided.

This pattern is likely driven by non-phonetic factors. For example, the Pinyin tone diacritic denoting T3, e.g., <mǎ>, implies that the ending of the tone should be high rather than low as shown in the ‘low’ target produced by the native speakers [12]. Moreover, the Tone 3 sandhi rule (/T3/→[T2]/_T3) may also contribute to the perceptual assimilation of the two tones. Beginner learners are introduced to this rule as early as in the second week due to common expressions such as /ni²¹⁴hao²¹⁴/ ([ni³⁵hao²¹⁽⁴⁾]) *you-good* ‘hello’. Along with the Pinyin T3 diacritic, the T2-T3 phonological alternation may mistakenly lead learners to assume that the f_0 rise at the end of T3 is indeed an indispensable part of this tone, which may make T2 and T3 even more perceptually similar than they actually are.

5. ACKNOWLEDGEMENTS

Thanks to the research assistants Yu-ming Chang, Sarang Jeong, Hyoju Kim and Felix Lin for help with data collection and labelling and the instructors of elementary Chinese courses for help with subject recruitment. This work was supported by MOST 107-2410-H-009-016-MY3 to PI Sang-Im Lee-Kim.

6. REFERENCES

- [1] Bassetti, B. 2006. Orthographic input and phonological representations in learners of Chinese as a foreign language. *Writ. Lang. and Lit.* 9(1), 95–114.
- [2] Huang, T., Johnson, K. 2010. Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonet.* 67. 243–267.
- [3] Kawahara, H., Masuda-Katsuse, I., de Cheveigné, A. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Com.* 27, 187–207.
- [4] Kuang, J. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *J. Acoust. Soc. Am.* 142(3), 1693–1706.
- [5] Mathieu, L. 2016. The influence of foreign scripts on the acquisition of a second language phonological contrast. *Sec. Lang. Res.* 32(2), 145–170.
- [6] Mok, P., Lee, A., Li, J. J., Xu, R. B. 2018. Orthographic effects on the perception and production of L2 mandarin tones. *Speech Com.* 101, 1–10.
- [7] Park, H., K., de Jong, J. 2008. Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *J. Phonet.* 36, 704–723.
- [8] Rochet, B. L., Fei, Y. 1991. Effect of Consonant and Vowel Context on Mandarin Chinese VOT: Production and Perception. *Can. Acoust.* 19(4), 105–106.
- [9] Schmidt, A. M. 1996. Cross-language identification of consonants. Part 1. Korean perception of English. *J. Acoust. Soc. Am.* 99(5), 3201–3211.
- [10] Showalter, C., Hayes-Harb, R. 2013. Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Sec. Lang. Res.* 29(2), 185–210.
- [11] Xu, C. X., Xu, Y. 2003. Effects of consonant aspiration on Mandarin tones. *J. Int. Phon. Assoc.* 33(2), 165–181.
- [12] Escudero, P., Hayes-Harb, R., Mitterer, H. 2008. Novel second-language words and asymmetric lexical access. *J. Phonet.* 36, 345–360.
- [13] Hayes-Harb, R., Cheng, H.-W. 2016. The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native English speakers. *Front. Psychol.* 7, 1–13.