MONOGRÁFICO

Círculo de Lingüística Aplicada a la Comunicación
ISSN: 1576-4737

http://dx.doi.org/10.5209/CLAC.65652

EDICIONES COMPLUTENSE

Challenges in multi-language pronunciation teaching: A cross-linguistic study of Chinese students' perception of voiced and voiceless stops

Liu Jiaqi 刘佳琦[1]; Zeng Ting 曾婷[2]; Lu Xiuchuan 鹿秀川[3]

Received: April 5, 2019 / Accepted: June 11, 2019

Abstract. This article reports on a cross-linguistic study of 58 Chinese students' perception of voiced and voiceless stops in their third language (L3). The participants were Japanese, Russian, or Spanish major students in a Chinese university, who were beginner learners of these languages but who had all learned English as their second language (L2) for over 10 years. The purpose of this study was to investigate the L3 learners' perceptual differences in the stop categories, and analyze the effects of the learners' multi-language background on their perception of L3 stops. Results from the perception experiment showed that: 1) the value and range of voice onset time (VOT) play an essential role in Chinese students' perception of voiceless stops; and 2) the pre-voicing during closure is the key to Chinese students' perception of voiced stops. We attribute their difficulty in perceiving L3 voiceless stops to the similarity in the phonemic range of voiceless stops between the learners' L3 and their L1 and L2, as this leads to confusion in perception. On the other hand, the dissimilarity between L3 voiced stops and those of L1 and L2 is conducive to the students' perception of L3 voiced stops. Findings from this study provide empirical evidence for the effect of similarity and dissimilarity in speech sounds as proposed in earlier phonology acquisition theories, and they can also inform the pedagogy of multi-language education.

Keywords: Multi-language education, voiced and voiceless stops, speech perception, VOT, Mandarin Chinese

[zh] 多语语音教育的挑战：中国学生清浊塞音感知习得的跨语言研究

摘要：本研究针对58名中国学生开展了第三语言（日语、俄语、西班牙语）清浊塞音感知习得的跨语言研究。被试是三语初级学习者，且拥有约10年的二语英语学习经历。本研究的目的是，调查中国学生的三语塞音种类感知差异，探明学习者的多语背景对三语塞音感知的影响。研究发现：(1）VOT值的大小和区间范围是中国学生感知清塞音的关键；(2）持阻阶段的带音性是中国学生感知浊塞音的关键。通过进一步讨论，我们发现三语清塞音感知之所以困难，是由于三语与母语、二语清塞音音位区间存在声学相似性，导致感知混淆。另一方面，三语浊塞音与母语、二语间存在的声学差异性则有利于三语浊塞音的感知。本研究成果不仅为跨语言视域下的第三语言语音习得理论发展提供了实证，而且为多语语音教学改革提供了依据。

关键词：多语教育，清浊塞音，语音感知，VOT，汉语普通话

[1] First author. Fudan University 复旦大学 (China) Email: jiaqiliu@fudan.edu.cn

[2] Corresponding author. Fudan University 复旦大学 (China); Shanghai Polytechnic University 上海第二工业大学 (China) Email: zengting@fudan.edu.cn

[3] Corresponding author. Fudan University 复旦大学 (China) Email: luxc@fudan.edu.cn

Contents.

How to cite: Liu, Jiaqi 刘佳琦; Zeng, Ting 曾婷 and Lu, Xiuchuan 鹿秀川 (2019). Challenges in multi-language pronunciation teaching: A cross-linguistic study of Chinese students' perception of voiced and voiceless stops. In: Yongyan Zheng 郑咏滟 and Xuesong (Andy) Gao 高雪松 (eds.) *Multilingual Research in the Chinese Context*. *Círculo de Lingüística Aplicada a la Comunicación* 79, 99-118 http://dx.doi.org/10.5209/CLAC.65652

1. Introduction

In the process of L3 acquisition, learners are generally influenced by their first language (L1) or second language (L2) (Williams & Hammarberg, 1998; Cenoz et al., 2001), and this influence is particularly obvious in the aspect of phonology (De Angelis, 2007; Onishi, 2016). The issues of native and non-native language transfer have received some attention in the area of L3 acquisition. In earlier studies L1 transfer was generally believed to be the most important factor affecting L3 acquisition, but there has been less importance attached to cross-linguistic influences from non-native languages (Ringbom, 1987; Pyun, 2005). In particular, the acquisition of L3 phonology is an emerging subdiscipline, and research in this area is limited in scope compared to studies on L3 lexis and morphosyntax (Hammarberg & Hammarberg, 2005; Llama et al., 2010).

We have learned from our language teaching experience that the acquisition of L3 voiced and voiceless stops is difficult for both teachers and learners in the initial stages of learning the language. Previous studies on the acquisition of stops have mostly been targeted at English language learners with various L1s (Flege, 1992). Earlier second language acquisition (SLA) literature on the acquisition of L2 stops has provided a lot of evidence for the transfer of L1 VOT (Lisker & Abramson, 1964; Abramson & Lisker, 1970; Abramson, 1977) values in the production of L2 stops (e.g. Flege, 1987; Flege & Hillenbrand, 1987). More advanced second language learners were found to be able to distinguish between L1 and L2 in terms of VOT, and their L2 VOT values were closer to those of native speakers (Caramazza et al., 1973; Flege, 1987, 1991). However, few studies to date have explored patterns in the acquisition of L3 stops (Tremblay, 2007; Llama et al., 2010; Wunder, 2010; Wrembel, 2014), so the breadth and depth of research in this area need to be expanded. The small range of languages studied in L3 phonology acquisition limits our understanding of how multilingual learners acquire different sounds during the initial stage of their language acquisition.

In the context of changes in China's foreign language education, however, findings from previous studies cannot fully explain and predict the issues arising in the acquisition of stops in this multi-language education environment. This study targeted Chinese university students majoring in Japanese, Russian, and Spanish, which are popular foreign languages alongside English, and conducted a cross-linguistic study of the learners' perception of voiced and voiceless stops. The study aims to compare the

Chinese L3 (Japanese, Russian, Spanish) learners' perceptual differences in the stop categories in order to determine the effects of the learners' multi-language background on their perception of L3 stops. The findings of this study will not only provide empirical evidence for the development of phonology acquisition theories; they may also be used to improve L3 phonology teaching in multi-language contexts.

2. Literature review

This section will review related studies from three aspects. First, we will compare and contrast the similarities and differences in word-initial stops in L1, L2, and L3. Second, we will summarize previous findings in the acquisition of stops. Third, the theoretical framework of this study will be explained.

2.1. Word-initial stops in L1, L2, and L3

Pre-voicing during stop closure and aspiration after release are two features that are generally used to distinguish among stop phonemes in languages. The participants in this study spoke Mandarin Chinese as their L1. Wu Zongji (1988) analyzed aspirated and unaspirated stops in Mandarin Chinese from physiological and acoustic perspectives. In particular, Wu Zongji pointed out that the intensity of airflow is crucial to distinguishing between aspirated and unaspirated stops for Chinese native speakers of Mandarin Chinese. The mean VOTs of the voiceless stops pronounced by Chinese native speakers are presented in Table 1 (adapted from Bao Huaiqiao and Lin Maocan (2014)). Table 1 shows that the VOT range for voiceless aspirated stops in Mandarin Chinese is 92.5ms to 102ms, and the VOT range for voiceless unaspirated stops is 6ms to 14.5ms.

Voicing	Aspiration	Place of Articulation	VOT
Voiceless	aspirated	Labial [p^h]	92.5ms
		Alveolar [t^h]	102ms
		Velar [k^h]	96.5ms
	unaspirated	Labial [p]	7.5ms
		Alveolar [t]	6ms
		Velar [k]	14.5ms

Table 1. Mean VOTs for Mandarin aspirated and unaspirated stops

English is a dominant foreign language among the languages taught in China (Chang, 2006), and it follows that students have usually already had experience in learning English as their L2 when they start learning their L3 at university. English is known to contrast voiced and voiceless phonemes in the word-initial position, but in fact voiced stops are regarded as having two possible phonetic realizations, i.e. voiced and voiceless unaspirated (Keating et al., 1983). The experiment reported by Lisker and Abramson (1964) showed that word-initial voiced stops in English are often not accompanied by pre-voicing—that is, there is no vibration of the vocal folds during the closure. Klatt (1975) and Docherty (1992) found that the word-initial voiced stops pronounced by native speakers of English did not have pre-voicing and the VOT values were positive (Table 2). Ladefoged and Keith (2011: 57) pointed out that the word-initial /p, t, k/ in English are voiceless aspirated, and the key difference between /p/ and /b/ is not whether it is voiced or not, but rather whether it is aspirated or not (i.e. the duration between the release and the onset of vocal fold vibration). The /p/ in English is a voiceless aspirated stop and /b/ is a voiceless unaspirated (perhaps voiced) stop. Whether /b, d, g/ are voiced or not mainly depends on their position in a word. Most native speakers of English produce no voicing during the closure when they pronounce the stops /b, d, g/ in the word-initial position, and the VOT values are positive.

	Klatt（1975）	**Docherty（1992）**
/p/	47ms	42ms
/t/	65ms	64ms
/k/	70ms	63ms
/b/	11ms	15ms
/d/	17ms	21ms
/g/	27ms	27ms

Table 2. Mean VOTs for English word-initial stops

The stops in Japanese, Russian, and Spanish are different from those in Chinese and English in the following ways. First, the word-initial stops in Japanese can be divided into two groups, i.e. voiceless aspirated [p^h, t^h, k^h] and voiced [b, d, g], according to the *Handbook of the International Phonetic Association* (International Phonetic Association, 1999: 117). Shimizu (1993) investigated the voicing feature of word-initial voiced and voiceless stops in six Asian languages and found that the mean VOTs for the word-initial stops [b] and [p^h] in Japanese were -89ms and 41ms respectively.

Second, Russian contrasts voiceless unaspirated stops and pre-voiced stops (Lisker & Abramson, 1964; Kulikov, 2012). Finally, the voiceless stops [p, t, k] in Spanish are followed immediately by vowels after the closure is released and the aspiration is short, while the voiced stops in Spanish are clearly pre-voiced (Abramson & Lisker, 1973; Ladefoged & Sandra, 2012).

To summarize, Table 3 shows the voiced and voiceless stop phonemes that distinguish meaning in the five languages involved in this study. Among the five languages, the word-initial stops in the learners' L1 (Chinese) and L2 (English) are mainly distinguished by aspiration, while the learners' L3s (Japanese, Russian, and Spanish) distinguish between voiced and voiceless stops according to whether there is any pre-voicing during the closure.

Language		**Voiceless**		**Voiced***
		Aspirated	Unaspirated	
L1	Mandarin Chinese	p^h、t^h、k^h	p、t、k	
L2	English	p^h、t^h、k^h		b、d、g
L3	Japanese	p^h、t^h、k^h		b、d、g
	Russian		p、t、k	b、d、g
	Spanish		p、t、k	b、d、g

* The phonemes of word-initial voiced stops in English are /b, d, g/, but there is no vibration of vocal folds during the closure for most native speakers of English and the VOT values are positive. There is vibration of vocal folds during the closure for the voiced stops in Japanese, Russian, and Spanish, and the VOT values are negative.

Table 3. Phonemes of word-initial stops in L1, L2, and L3

2.2. Acquisition of stops

Among researchers who have investigated the acquisition of stops, Flege (1992) carried out a wide range of empirical studies involving English learners of various L1s. Flege (1992: 577) found that the stops in the learners' first language are closely related to the acquisition of stops in English. If the voiceless stops in the learners' L1 have a relatively short lag VOT, the voiceless stops in the learners' English interlanguage will also have a short lag VOT, or a VOT that is somewhere between L1 and English. In terms of perception, Bohn and Flege (1993) identified that the outcome of the early stage of L2 learning for Spanish English learners will be influenced by the initial 'mapping' of L2 sounds onto L1 categories via interlingual identification, rather than determined by the VOT values of the stops.

Among studies that focused on Chinese learners, Flege (1992: 568) demonstrated that Chinese learners of English could not make an acoustic distinction between /p/ and /b/ in the early stage of acquisition, and could not discriminate between voiced and voiceless stops in English. Fukuoka (1995) and Liu Jiaqi (2011) conducted empirical studies of Chinese learners' perception of voiced and voiceless stops in Japanese and showed that the learners were confused about the difference between voiced and voiceless stops, and it was especially difficult for the learners to perceive and produce intervocalic stops. Liu Jiaqi (2011) argued that the difficulty in distinguishing between voiced and voiceless stops might be attributed to a combined effect of the difference and similarity in speech sounds between Chinese and Japanese.

In the area of third language acquisition (TLA), some VOT patterns have been found in the acquisition of stops (Tremblay, 2007; Llama et al., 2010; Wrembel, 2014). In particular, Wrembel (2014) investigated cross-linguistic influence in the acquisition of third language phonology by exploring the VOT patterns in L3 learners' production as a result of the interaction between three phonological systems. Wrembel carried out two parallel studies involving different language combinations: (1) L1 Polish, L2 English, and L3 French; and (2) L1 Polish, L2 English, and L3 German. She compared the multilingual subjects' VOT values for the three phonological systems, and revealed unique L3 interlanguage VOT patterns. There were notable differences in voiceless stops between the participants' L1 and L2, and the VOT values of their L3 production were intermediate between the mean L1 and L2 VOTs, probably under the influence of the voiceless stops with long-lag VOT in L2 English. The results to some extent supported the study's hypothesis about a combined cross-linguistic influence from L1 and L2 on L3 acquisition.

The research also found that when learning three or more languages, learners demonstrated increased metalinguistic awareness as they actively built a new phonetic system for the additional foreign language to distinguish it from their native language and other already acquired foreign languages. However, the findings of the research were limited as the author only explored the VOT patterns for voiceless stops and did not fully analyze the patterns in the acquisition of a phonological system that contrasts voiced and voiceless stops for L1, L2, and L3, along with the associated cross-linguistic influence. In addition, the research only focused on learners' production of stops and did not investigate their perception of stops.

2.3. Theoretical framework

Regardless of L2 or L3 acquisition, the differences and similarities between languages are always two key factors when we predict or analyze the features of acquisition. Earlier research holds that the degree of difficulty in acquisition depends on the difference between L1 and the target language—the more differences, the more difficulties. Lado (1957) proposed the Contrastive Analysis Hypothesis (CAH), which predicts and accounts for phenomena in language acquisition by comparing the differences between L1 and the target language. Eckman (1977) put forward the Markedness Differential Hypothesis (MDH) on the basis of the CAH. The MDH draws on the markedness feature to account for the different levels of difficulty in language acquisition. Eckman (1991) further improved the MDH and put forward the Interlanguage Structural Conformity Hypothesis (ISCH). These hypotheses all center around the notion of differences between languages. When a feature in the target language is different from the L1, it will be difficult for learners to acquire it. However, research on language acquisition shows that some issues in language acquisition cannot be predicted or explained by linguistic typological differences.

Oller and Ziahosseiny (1970) discussed the effect of the similarity between L1 and L2 on language acquisition. Their study showed that it is easier to acquire a linguistic feature that is different from the learner's L1 than a feature that is similar to the learner's L1. Flege (1987) put forward the notion of 'equivalence classification'. He held that similar linguistic features are difficult to learn because the learners classify and perceive these features to be similar to those in their L1, which will lead to stagnancy in acquisition. By contrast, features that are new or significantly different from those in the learners' L1 are more easily noticed by the learners and therefore they can be learned earlier. Flege's (1995) Speech Learning Model (SLM) concluded that '[t]he greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned' (p. 239). According to the SLM, L2 learners are likely to create an interlanguage category by 'compromising' or 'hybridizing' the VOT values for both languages, which may deviate from both L1 and L2, in order to maintain the phonetic contrast between the two languages.

On the basis of Flege's research findings, Major and Kim (1999) proposed the Similarity Different Rate Hypothesis (SDRH) and further discussed the effect of similarity and difference on language acquisition. This hypothesis claims that dissimilar phenomena are acquired at a faster rate than similar phenomena, and that markedness is a mediating factor that slows the rate of acquisition. In other words, it is the similarity between languages, rather than the difference between them, that has a negative effect on language acquisition. Chan (2012, 2014) investigated Cantonese-speaking ESL learners' perception of English speech sounds. This study showed that Cantonese-speaking ESL learners' perception of English speech sounds was intimately related to their perceived similarity between certain L1 and L2 contrasts, which reaffirmed the adverse effects of similarity on L2 phonology acquisition. The above studies all point to the conclusion that similarity between L1 and L2 will adversely affect language acquisition.

Building on these findings, Ringbom and Jarvis (2009: 106) further demonstrated the important effect of the similarities between languages from the perspective of multilanguage acquisition and teaching. They proposed that for learners, similarities have a much more direct effect on language learning and performance than differences do. Learners are constantly trying to establish links between the target language and whatever language they have acquired in the past. Learners tend to look for similarities rather than differences. Ringbom and Jarvis also pointed out the need to distinguish between actual similarities and assumed similarities. Actual similarities can be analyzed linguistically, while assumed similarities only take place in the learner's mind. Learners will establish over-simplified cross-linguistic mapping between languages (i.e. assuming similarities between languages) in order to reduce the workload, which is often not beneficial to acquisition. It is necessary for teachers to encourage learners to make use of actual similarities to prevent their exaggerated reliance on merely assumed similarities.

To summarize the reviewed studies, earlier research on the acquisition of stops mainly focuses on English language learners with different L1s. There are acoustic experiments that have supported the existence of cross-linguistic influence in the acquisition of stops, but the findings from these studies are still limited. The current study was conducted in the context of foreign language education in Chinese universities, where L3 teachers and learners are in a complex multi-language environment. L3 learners generally have over 10 years of experience learning English. This study explored the patterns in L1 Mandarin students' perception of different categories of stops in L3 and the reasons behind the patterns. The study also focused on the differences and similarities among the learners' L1, L2, and L3 in terms of the distinction between voiceless and voiced stops, in order to uncover the cross-linguistic influence that is affecting learners' perception of different categories of stops in L3.

The study addressed the following two research questions:

(1) Will learners' L1 and L2 phonological systems affect their perception of L3 voiceless stops, and what will be the effects of this?

(2) Will learners' L1 and L2 phonological systems affect their perception of L3 voiced stops, and what will be the effects of this?

3. The study

3.1. Participants

The perception experiment recruited 58 participants, specifically 46 females and 12 males aged between 18 and 20 years old. None of the participants reported any history of speech and hearing disorders. The participants of this experiment were divided into the following three groups: 1) L1 Mandarin Chinese, L2 English, L3 Japanese (*n* =20); 2) L1 Mandarin Chinese, L2 English, L3 Russian (*n* =19); 3) L1 Mandarin Chinese, L2 English, L3 Spanish (*n* =19). They were all students enrolled in their respective language major programs in a Chinese university. They only started learning their L3 language after they entered the university, and had been learning the language for two months by the time of the perception experiment. They had just learned the pronunciation and spelling of the target language, and were all beginner-level learners. This study controlled for the participants' birth place and the language they spoke at home. All the participants were born in northern or northwestern China; the dialects in these areas are similar to Mandarin Chinese in terms of the stop consonants. Each participant had been learning English for more than 10 years at the time of the experiment.

The phonetic stimuli in this experiment were provided by two native speakers of different genders for each of the three languages under investigation (i.e. Japanese, Russian, and Spanish)—altogether six contributors. The age of these stimuli providers ranged from 25 to 45 years. They were either foreign students studying in China or foreign language teachers. The two Japanese native speakers grew up in Tokyo and Osaka respectively; both Russian native speakers grew up in Moscow; and the two Spanish native speakers grew up in Segovia and Málaga respectively.

3.2. Materials

There were 24 stimuli and 12 fillers for each of the three tested languages in this experiment. These stimuli were read by two native speakers of each language (a male and a female), so altogether there were 48 stimuli for each language. The stimuli were bilabial /p/ and /b/, alveolar /t/ and /d/, and velar /k/ and /g/, all in the initial position of monosyllabic or disyllabic words. The stops were all followed by the vowel [a]. There was no stressed syllabic in the stimuli. Refer to the Appendix for the stimuli used in this experiment.

3.3. Procedure

3.3.1. Stimuli preparation

Two native speakers of opposite sexes provided the stimuli for each of the three tested languages. The 48 stimuli were inserted into sentences in the same language. For example, 'これは______' (Japanese), 'Вот______' (Russian), and 'Es______' (Spanish). Each sentence was read three times by the native speakers, and their second production was used as the stimulus for the experiment. The recording was conducted in a quiet sound-proof environment using a TASCAM DR-44WL linear PCM recorder (44.1 kHz sampling rate, 16-bit quantization level) and a AKG

C544L head microphone. The audio files were saved in .wav format and were later extracted for acoustic analyses.

This study used Praat 6.0 to create the phonetic stimuli. Each stimulus was preceded by a 400ms period of silence to leave a gap between stimuli. A synthetic sound with a fundamental frequency of 500 Hz and a duration of 400ms was added as a start signal to alert the participants. Finally, a 1000ms period of silence was added to the end of each stimuli so that the participants would have enough time to read and understand the choices on the computer screen.

3.3.2. Perception test

The perception test for this study used a Praat 6.0 ExperimentMFC 6 script. Laptops and SONY MDR-ZX110NC noise-cancelling headphones were used in the test. During the test the participants, wearing headphones, faced a computer screen in a quiet environment. Minimal pairs of phonetic stimuli randomized by ExperimentMFC 6 were shown on the computer screen. The participants used a mouse to select the stimuli they heard.

Before the participants started the perception test they were asked to complete a questionnaire about their first language background and foreign language learning experience, and do a mini perception test to familiarize them with the experimental equipment and procedure. The mini test was similar to the real test in procedure, but not relevant in content. Each participant spent 10 to 15 minutes doing the real test. They were allowed to have a break after every five stimuli. When the test ended, Praat automatically recorded the participants' perception accuracy and reaction times. The researchers saved the Praat file as an Excel document for later analysis.

3.4. Analysis

Various studies in acoustic phonetics and perceptual phonetics have demonstrated that whether a sound is voiced or voiceless is related to many acoustic features—for example, pre-voicing, aspiration (Klatt, 1975), closure duration, energy density, F1 cutback, F1 onset formant and transitions, F0 adjacent to the closure and F0 contour, duration of the preceding vowel, F1 adjacent to closure, amplitude of F1 at release, and so on. Among these features, pre-voicing and aspiration are directly related to the interval between release and the onset of vocal fold vibration, and therefore they are important features in determining whether a sound is voiced or voiceless. As these two features are captured by the measure of VOT (Lisker & Abramson, 1964; Abramson & Lisker, 1970; Abramson, 1977), this study mainly focused on and compared the VOT of voiced and voiceless stops in L1, L2, and L3 and the relation between VOT and perception accuracy.

SPPAS (Ver.1.8.6) was used to annotate the speech stimuli, and the software's automatic annotation was checked and confirmed by the researchers. The annotated acoustic parameters were extracted on different tiers using the 'analyse_tier.praat' script created by Daniel Hirst (Ver.2012/10/30). The researchers used R(Ver.3.4.0) to clean the data, conduct statistical analyses, and generate figures.

4. Results

The researchers recorded the VOT values of L2 English stops for the three groups of L3 learners and compared them with the VOT values in Mandarin Chinese. The accuracy of the three groups of L3 learners' perception of voiceless and voiced stops was analyzed statistically. The results will be presented in this section.

4.1. Acquisition of L2 stops

To determine the learners' acquisition of L2 stops, the learners were asked to undertake a production test of voiced and voiceless stops in L2 English. The recording procedures and test equipment used in this test were the same as those described in Section 3.3.1. The items in this test were monosyllabic words with /p, t, k/ and /b, d, g/ in the initial position—for example, *park*, *bark*, etc. The acoustic features of the data were annotated and the VOT values were extracted from the data. A one-way ANOVA analysis showed that there was no significant difference in the VOT values of L2 English voiced and voiceless stops among the three groups of learners (voiceless stops: $F(2, 171)=1.769$, $p=.174$; voiced stops: $F(2, 171)=2.908$, $p=.057$). This means that the acquisition of L2 English stops was at a similar level for the three groups of learners before the experiment.

The distribution of the VOTs for the L2 English voiced and voiceless stops pronounced by Chinese learners is shown in Figure 1 below. As can be seen from the figure, the interquartile of the L2 word-initial voiceless stops /p, t, k/ is 65–95ms and the interquartile of the voiced stops /b, d, g/ is 10–25ms. All the VOTs for the L2 word-initial stops are greater than 0. The major difference between voiced and voiceless stops is the value of VOT. The threshold VOT to separate L2 voiced and voiceless stops, as shown in the learners' production, is about 30–35ms.

To determine the relationship between the participants' L1 and L2 with respect to the VOT values of voiced and voiceless stops, tdhe VOT values for stops in L2 English (Figure 1) and the benchmark VOT values in Mandarin Chinese (as shown in Table 1) were compared. It was found that the learners' range of VOTs for L2 voiceless stops was approximately the same as that in L1, and the range of VOTs for L2 voiced stops was also similar to that for L1 voiceless unaspirated stops.

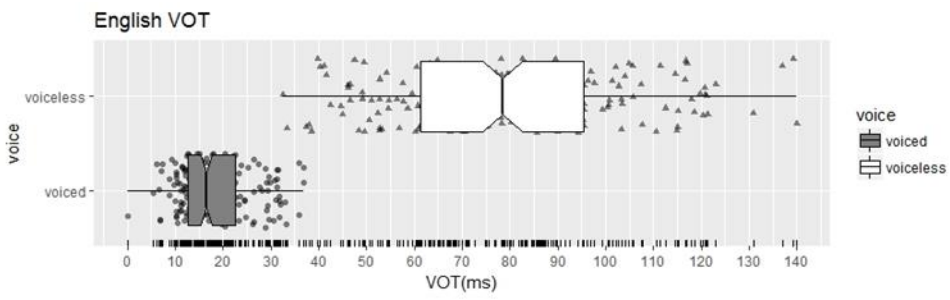
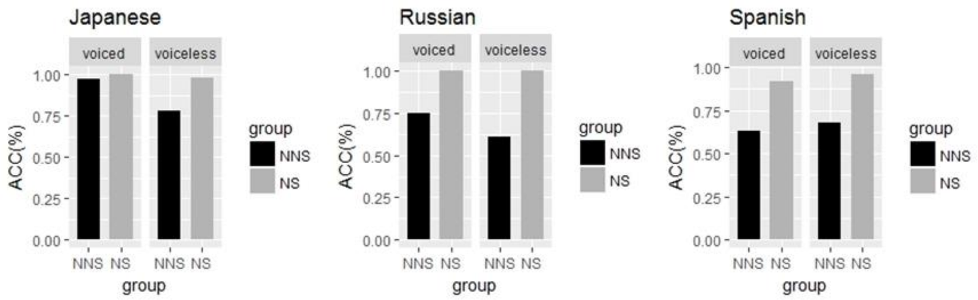


Figure 1. Distribution of VOTs for L2 voiced and voiceless stops

4.2. Results for the perception of L3 stops

The results of the experimental testing of L1 Chinese students' perception of L3 voiced and voiceless stops showed that non-native speakers (NNS) had a lower accuracy (ACC) than native speakers (NS) (see Figure 2). The results of an independent-sample *t*-test showed that NNSs' accuracy was significantly lower than NSs' accuracy in perceiving stops for all three L3s. The *t* statistics and *p* values were: $t(146)=7.038$, $p<.001$ for Japanese voiceless stops; $t(479)=3.79$, $p<.001$ for Japanese voiced stops; $t(455)=16.99$, $p<.001$ for Russian voiceless stops; t(455)=12.32, p<.001 for Russian voiced stops; t(111)=7.64, p<.001 for Spanish voiceless stops; and t(80)=6.17, p<.001 for Spanish voiced stops.



.

Figure 2. Mean accuracy in perceiving L3 voiced and voiceless stops

Figure 2 displays the mean accuracy in perceiving L3 voiced and voiceless stops for each language. The accuracy in perceiving voiced stops is higher than voiceless stops for both Japanese ($t(479)=-9.194$, $p<.001$) and Russian ($t(455)=-4.34$, $p<.001$). There was no significant difference in accuracy in perceiving the voiced and voiceless stops for Spanish ($t(455)=1.45$, $p=.147$).

4.2.1. L3 voiceless stops

The perception test showed that NNSs' accuracy in perceiving voiceless stops was lower than NSs' accuracy. It is worth noting that there was a significant difference in the accuracy of perceiving voiceless stops among the three groups of L3 learners, as shown by a one-way ANOVA ($F(2, 69)=3.676$, $p<.05$). A TukeyHSD post-hoc test revealed that Japanese learners' accuracy for voiceless stops was significantly higher than that of Russian learners ($p<.05$), but there was no significant difference between Russian and Spanish learners ($p=.25$) or between Japanese and Spanish learners ($p=.53$).

Pearson's correlation test was used to estimate the strength of the correlation between learners' accuracy perceiving L3 voiceless stops and the VOTs of the stimuli (see Figure 3). The correlation test showed that: 1) Japanese learners' accuracy was correlated with the VOTs of the stimuli ($r=.41$); the closer the VOT was to zero, the lower the accuracy in perceiving voiceless stops, and therefore the value of VOT is expected to provide a cue for the voicing of stops for Japanese learners; and 2) there was no significant correlation between accuracy and VOTs for either Russian learners (*r*=.32) or Spanish learners (*r*=.20), suggesting that the VOTs of voiceless stops in Russian and Spanish are not helpful for Chinese learners to determine whether a stop is voiced or voiceless.

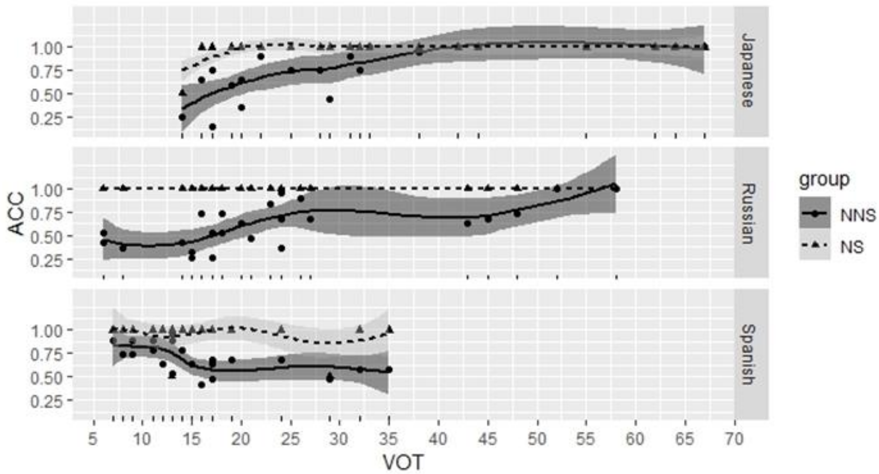


Figure 3. Correlation between VOTs of voiceless stops and perception accuracy for L3 NNSs and NSs

4.2.2. L3 voiced stops

It was shown in Figure 2 that: 1) accuracy in perceiving voiced stops was much higher than accuracy in perceiving voiceless stops for both Japanese and Russian; and 2) there was no significant difference in accuracy between voiced and voiceless stops for Spanish. Acoustic analysis showed that the VOTs of voiced stops pronounced by native speakers in the stimuli were all below zero. The mean VOTs were -74.42ms, -95.46ms, and -88.25ms for Japanese, Russian, and Spanish respectively. The learners' VOTs for L1 voiceless unaspirated stops and L2 voiced stops were all above zero, which is markedly different from the VOTs of L3 voiced stops, but Chinese learners' accuracy in perceiving L3 voiced stops was higher than or the same as their accuracy for voiceless stops. In other words, for L1 Chinese learners, the perception of voiced stops in their L3 was not more difficult than voiceless stops.

The correlation between learners' accuracy in perceiving L3 voiced stops and the VOTs of the stimuli was investigated. Results showed that learners' accuracy in perceiving L3 voiced stops was not significantly correlated with the VOTs of the stimuli for all three languages (Japanese, *r* =.02; Russian, *r* =.16; Spanish, *r* =.01). In other words, the value of VOT was not an effective cue for L3 voiced stops for L1 Chinese learners.

The findings from this study indicate that learners' perception of L3 voiced stops was not based on the value of VOT, but rather on whether the VOT value was positive or negative. In other words, whether L3 voiced stops have a pre-voicing feature is the key to Chinese learners' perception of voiced stops.

5. Discussion

5.1 Perception of L3 voiceless stops and its cross-linguistic influence

The perception experiment showed that learners' accuracy in perceiving voiceless stops was different among the three groups of L3 learners. Japanese learners' accuracy in perceiving voiceless stops was higher than that of Russian learners. Japanese learners' accuracy was also correlated with the VOTs of the stimuli; the shorter the VOT, the lower the accuracy in perceiving the voiceless stops. However, there was no significant correlation between accuracy and VOTs for Russian and Spanish learners.

This finding can probably be explained by the distribution of VOTs for the voiceless stop stimuli in the experiment (see Figure 4). As shown in Figure 4, there was a clear difference in the distribution of VOTs among the three L3s. The mean VOT of voiceless stops in Japanese was relatively long at 33ms, with a median of 30ms and a fairly large standard deviation (SD) of 15.71. By comparison, the median of the VOTs of voiceless stops in Russian (20.5ms) and Spanish (15ms) were both shorter than in Japanese. Russian and Spanish also both had a smaller range, below 30ms, and a smaller spread, with SDs of 14.22 and 7.33 respectively. A VOT of 30ms is a universally meaningful phonetic threshold. It has been observed that many languages have a boundary between aspirated and unaspirated stops at 30ms VOT (as cited in Keith, 2003: 101). This cross-linguistic phonetic property is constrained by the auditory system of human beings. Our finding has confirmed this hypothesis.

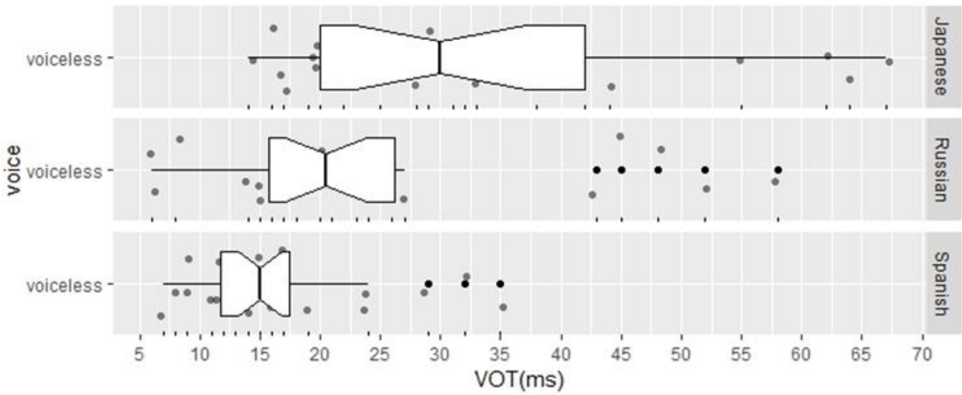


Figure 4. Distribution of VOTs of voiceless stops in the three L3s

In the long-term process of language acquisition, learners may have to make a compromise between the native and target language patterns of phonetic implementation, and such a compromise may 'reflect a restructuring of the phonetic space so that it encompasses both languages' (Flege, 1981: 451). Bohn and Flege (1993) discovered that in the early stage of acquisition, Spanish English learners' accuracy in perceiving word-initial stops was related to the mapping between L1 and L2 stop categories. Learners would make use of the phonological features of their L1 or L2 to learn L3 phonology. It is therefore expected that in the early stages of L3 acquisition learners are likely to compare the phonological features of L3 with those of L1 or L2, and learn new speech sounds on the basis of earlier learned phonological systems. The VOT range of Chinese learners' L1 voiceless aspirated stops is 92.5–102ms (see Table 1) and the VOT range of their L2 voiceless aspirated stops is 65–95ms (see Figure 1); both ranges are considerably longer than 30ms. Among the three L3s for the three groups of learners, with a mean of 33ms and a median of 30ms, the VOT values of Japanese voiceless stops are greater than those of Russian and Spanish. By comparison, Japanese voiceless stops have a similar VOT range to the voiceless stops in the learners' L1 and L2, which makes it easier for the learners to notice the aspiration of Japanese voiceless stops and to separate voiced and voiceless stops—hence the higher perceptual accuracy.

Compared to L3 Japanese, the learners' perception of L3 Russian and Spanish voiceless stops was lower. Our analysis showed that the VOT range of Chinese learners' L1 voiceless unaspirated stops was 6–14.5ms (Table 1) and the VOT range of their L2 voiced stops was 10–25ms (Figure 1). The VOT range of voiceless stops in L3 Russian and Spanish is very close to the VOT range of learners' L1 unaspirated stops and L2 voiced stops. Oller and Ziahosseiny (1970), Flege (1987), Flege (1995), and Major and Kim (1999) all pointed out the adverse effect of similarity between L1 and foreign language on language acquisition. The results of this study showed that when the VOT ranges of L1 voiceless unaspirated stops and L2 voiced stops are similar to the VOT range of L3 voiceless stops, learners are likely to experience perceptual confusion about phonemes and show poor performance.

Ringbom and Jarvis (2009: 106) noted that similarities between languages have a direct effect on language acquisition. In the early stage of acquisition in particular, learners are likely to establish over-simplified cross-linguistic mapping based on assumed similarities between languages, which will often adversely affect language acquisition. In the current study, the VOT range of voiceless stops in L3 Russian and Spanish was very close to the VOT range of learners' L1 unaspirated stops and L2 voiced stops. Learners are very likely to establish over-simplified mapping between these phonemes in their mind, and as a result experience perceptual confusion.

The findings from this study also suggest that this hypothesis does not only apply to L2 acquisition but also to Chinese learners' acquisition of L3 stops. When L1 and L2 have a similar stop system, this may affect L3 acquisition of stops during the same process.

To summarize, evidence suggests that: 1) learners' accuracy in perceiving L3 voiceless stops is closely related to the range and spread of VOT values. Japanese voiceless stops have a relatively large mean VOT and spread, so Chinese learners were able to use the VOT value to help distinguish between voiced and voiceless stops, while Russian and Spanish voiceless stops have a VOT range that is smaller than 30ms and a small spread and therefore cannot be utilized as an effective cue to the voicing feature for Chinese learners; and 2) in the early stage of L3 phonology acquisition, learners have not yet established the L3 phonemic categories in perception and therefore are very likely to make use of over-simplified cross-linguistic mapping to perceive the phonemes. When the acoustic parameter representing different phonemes is similar among L1, L2, and L3, learners tend to experience perceptual confusion about phonemes and show poor performance.

5.2 Perception of L3 voiced stops and its cross-linguistic influence

The results of the experiment showed that although the VOT values of learners' L1 unaspirated stops and L2 word-initial voiced stops were all above zero, which is different from the acoustic features of L3 voiced stops, Chinese learners' accuracy in perceiving L3 voiced stops turned out to be higher than or equal to their accuracy for voiceless stops. Earlier second language acquisition research often attributed the difficulty in acquisition to the difference between L1 and the target language (Lado, 1957; Eckman, 1977; Eckman, 1991). However, the 'equivalence classification' hypothesis (Flege, 1987) and SLM (Flege, 1995) suggested that sounds in the target language that are significantly similar to those in L1 are harder to acquire because learners classify or perceive them to be equivalent to those in their L1, while new or slightly similar sounds are easier to learn because learners are more likely to notice the difference between L1 and the target language.

It is widely recognized that although VOT is a continuum, it is constrained by the categorical perception of human beings (Liberman et al., 1957) and learners are far more capable of distinguishing between stop categories than between vague acoustic features. If learners can detect the phonetic difference between an L3 sound and an L1 or L2 sound and categorize them as two different sounds, it will be easier for them to establish a phonetic category in L3. On the other hand, if learners can barely distinguish a very similar target language sound from a sound in L1 or L2, they will not continue to verify the phonetic difference between languages and reproduce it, which will cause difficulty in the learners' perception (Kingston, 2003). The biggest difference in the voiced stop system between L3 and L1 or L2 lies in the feature of pre-voicing. When learners notice the presence or absence of pre-voicing, they can readily distinguish between voiced and voiceless stops.

Findings about the VOT patterns in learners' L3 acquisition of voiceless stops from Wrembel's two studies (2014) have to some extent substantiated the assumption of a combined cross-linguistic influence in L3 acquisition. Wrembel's research also found that when learners were learning three or more languages, they demonstrated increased metalinguistic awareness. They were actively building a new phonetic system for the additional foreign language to distinguish it from their native language or other already acquired foreign languages. Unfortunately, the studies only investigated the VOT patterns for voiceless stops and did not fully analyze the patterns in the acquisition of a phonological system that contrasts voiced and voiceless stops for L1, L2, and L3 and the associated cross-linguistic influence. The current study found similar patterns in the perception of L3 voiced stops, suggesting that multilanguage learners can better perceive the features of L3 voiced stops through comparing L3 with L1 or other already acquired foreign languages in order to establish a new phonemic system for the additional language. This finding is in line with Flege's (1995) Speech Learning Model (SLM), which claims, among other things, that the phonetic system of a learner remains adaptive throughout their lifetime and is open to modifications of phonetic categories.

To recapitulate the above discussion, learners' accuracy in perceiving L3 voiced stops was higher than or equal to their accuracy for voiceless stops. This finding indicates that learners are more sensitive to the difference in voiced stops between L3 and L1 or L2 than they are to the difference in voiceless stops. In other words, if the word-initial stops in learners' L1 and L2 basically have no pre-voicing feature, the difference in L3 word-initial voiced stops is more salient. Therefore, learners are more aware of it and perceptual acquisition is facilitated.

6. Conclusion

This study was conducted in the context of China's unprecedented and complex multi-language education environment. The study investigated L1 Mandarin Chinese learners' perception of L3 word-initial stops, and explored the cross-linguistic influence of learners' L1 and L2 stop systems on L3 acquisition. It was found that: 1) learners' perception of L3 voiceless stops was related to the VOT range of the stimuli, while learners' perception of L3 voiced stops mainly depended on the presence or absence of the pre-voicing feature; and 2) the difficulty in acquiring L3 voiceless stops was attributed to the similarity in phonemic range between L3 and L1 or L2. Learners' accuracy in perceiving L3 voiced stops was higher than or equal to their accuracy for voiceless stops, which can be explained by the significant differences between L3 and L1 or L2. The findings demonstrate that it is similarities between L3 and earlier learned languages, rather than the differences between them, that cause the main difficulty in acquiring L3 stops.

The results of this study provide strong support for the SLM (Flege, 1995) and the SDRH (Major & Kim, 1999). The results also suggest that the challenge of phonology to multi-language education comes not only from the target language itself, but is also closely related to the difference and similarity between L3 and L1 or L2, and this directly impacts on the outcome of multi-language education. On the basis of the findings of this study, we make the following suggestions regarding phonology teaching in the multi-language context.

First, teachers should probably be acquainted with the learners' L1 and L2 in addition to the phonology features of the target language, and should understand the differences and similarities between these languages. Second, it is advisable that teachers make use of differences and similarities between languages to predict issues that might arise in acquisition and improve pedagogy. Third, teachers should explicitly point out actual similarities and encourage learners to make use of them, to prevent learners from over-relying on merely assumed similarities.

There are some limitations with the current study. Learners' perception of L3 stops will change as they spend more time learning the language. For example, language attrition, learning plateau, or even degeneration may occur. It would be advisable to conduct a longitudinal study in order to determine how learners restructure their phonological space in the process of acquisition. In addition, the data from the current study cannot address how L1 and L2 affect L3 acquisition when L1 and L2 have different phonological features, and therefore further experiments are needed to investigate this question.

Acknowledgement

This work was supported by the National Social Science Foundation of China [Grant Number 18BYY227].

References

Abramson, A. S. 1977. Laryngeal timing in consonant distinctions. *Phonetica.* 34, 295–303, http://www.haskins.yale.edu/sr/SR047/sr047_08.pdf.

Abramson, A. S., & Lisker, L. 1970. Laryngeal behavior, the speech signal and phonological simplicity. *Actes du xe Congres International des Linguistes.* 4, 123–129, http://www.haskins.yale.edu/Reprints/HL0107.pdf.

Abramson, A. S., & Lisker, L. 1973. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics.* 1, 1–8.

Bao, Huaiqiao 鲍怀翘, & Lin, Maocan 林茂灿. 2014. *实验语音学概要 [An Introduction to Experimental Phonetics]*. Peking: 北京大学出版社 Peking University Press.

Bigi, B. 2017. *SPPAS: The automatic annotation and analyses of speech (Version 1.8.6) [Computer software]*, http://www.sppas.org.

Boersma, P., & Weenink, D. 2009. *Praat: Doing phonetics by computer (Version 6.0) [Computer program]*, http://www.fon.hum.uva.nl/praat.

Bohn, O. S., & Flege, J. E. 1993. Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics.* 21, 267–290.

Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*. 54, 421–428, DOI: 10.1121/1.1913594.

Cenoz, J., Hufeisen, B., & Jessner, U. 2001. *Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. In the series Bilingual Education and Bilingualism, 31. Clevedon: Multilingual Matters.

Chan, A. Y. W. 2012. Cantonese English as a second language learners' perceived relations between "Similar" L1 and L2 speech sounds: A test of the Speech Learning Model. *The Modern Language Journal.* 96 (1), 1–19, DOI: 10.1111/j.1540-4781.2012.01291.

Chan, A. Y. W. 2014. The perception and production of English speech sounds by Cantonese ESL learners in Hong Kong. *Linguistics.* 52 (1), 35–72, DOI: 10.1515/ling-2013-0056.

Chang, J. 2006. Globalization and English in Chinese higher education. *World Englishes.* 25 (3–4), 513–525, DOI: 10.1111/j.1467-971X.2006.00484.

De Angelis, G. 2007. *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.

Docherty, G. J. 1992. *The Timing of Voicing in British English Obstruents*. Berlin: Foris Publications.

Eckman, F. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning.* 27, 315–330, DOI: 10.1111/j.1467-1770.1977.tb00124.x.

Eckman, F. 1991. The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition.* 13, 23–41, DOI: 10.1017/S0272263100009700.

Flege, J. E. 1981. The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly.* 15, 443–455, DOI: 10.2307/3586485.

Flege, J. E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics.* 15, 47–65.

Flege, J. E. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America.* 89, 395–411, DOI: 10.1121/1.400473.

Flege, J. E. 1992. Speech learning in a second language. In C. Ferguson, & L. Menn (Eds.), *Phonological Development: Models, Research, and Application* (pp. 565–604). Timonium, MD: York Press.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (Eds.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 233–277). Baltimore: York Press.

Flege, J. E., & Hillenbrand, J. 1987. Differential use of closure voicing and release burst as cue to stop voicing by native speakers of French and English. *Journal of Phonetics*. 15, 203–208.

Fukuoka, M. 1995. Pekingo/syanhaigo wo bogo to suru nihongo gakusyusya no yusei/musei haretsuon no oudanteki oyobi zyudanteki syutoku kenkyu [A longitudinal and cross-sectional study of the acquisition of Japanese voiced-voiceless plosives by speakers of Peking and Shanghai dialects]. *Nihongo Kyouiku* [Journal of Japanese Language Teaching]. 87, 40–53.

Hammarberg, B., & Hammarberg, B. 2005. Re-setting the basis of articulation in the acquisition of new languages: A third-language case study. In B. Hufeisen, & R. Fouser (Eds.), *Introductory Readings in L3* (pp. 11–18). Tübingen: Stauffenburg Verlag.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Keating, P. A., Linker, W., & Huffman, M. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics.* 11, 277–290.

Keith, J. 2003. *Acoustic and Auditory Phonetics (Second Edition)*. Malden, MA: Blackwell.

Kingston, J. 2003. Learning foreign vowels. *Language and Speech.* 46 (2–3), 295–349, DOI: 10.1177/00238309030460020201.

Klatt, D. H. 1975. Voice Onset Time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research.* 18, 686–706, DOI: 10.1044/jshr.1804.686.

Kulikov, V. 2012. *Voicing and Voice Assimilation in Russian Stops*. Unpublished Doctoral dissertation, University of Iowa, https://ir.uiowa.edu/etd/3327.

Ladefoged, P., & Sandra, F. D. 2012. *Vowels and Consonants (Third Edition)*. West Sussex: Wiley-Blackwell.

Ladefoged, P., & Keith, J. 2011. *A Course in Phonetics (Sixth Edition).* Boston: Wadsworth.

Lado, R. 1957. *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.

Liberman, A. M., Harris, K. S., & Hoffmand, H. S. 1957. The discrimination of speech sounds within and across phonetic boundaries. *Journal of Experimental Psychology.* 54, 358–368.

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

Liu, Jiaqi 劉佳琦. 2005. Chugoku hogen washa niyoru nihongo yusei/musei haretsuon no chikaku ni kansuru ichikosatsu [A study of the stop voicing contrast by Chinese learners of Japanese: Based on dialect of China]. *Waseda daigaku Nihongo kyoiku kenkyu* [Waseda Journal of Japanese Applied Linguistics]. 6, 79–89, http://hdl.handle.net/2065/3550.

Liu, Jiaqi 劉佳琦. 2011. *Nihongo Yusei/musei Haretsuon no Syutoku Oyobi Kyouiku* [The Acquisition and Education of Japanese Voiced and Voiceless Stop Consonants]. Seoul: Shinsung Publisher.

Liu, Jiaqi 劉佳琦. 2019. Daisan gengo toshiteno nihongo haretsuon no chikaku syutoku nitsuite [The perceptual acquisition of L3 Japanese stop contrasts by Mandarin Chinese speakers]. *Waseda Nihongo Kyoikugaku* [Waseda Studies in Japanese Language Education]. 26, 127–146, http://hdl.handle.net/2065/00062886.

Llama, R., Cardoso, W., & Collins, L. 2010. The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism.* 7 (1), 39–57, DOI: 10.1080/14790710902972255.

Major, R. C., & Kim, E. 1999. The similarity differential rate hypothesis. *Language Learning.* 49, 151–183, DOI: 10.1111/0023-8333.49.s1.5.

Oller, J. W., & Ziahosseiny, S. M. 1970. The contrastive analysis hypothesis and spelling errors. *Language Learning.* 20, 183–189.

Onishi, H. 2016. The effects of L2 experience on L3 perception. *International Journal of Multilingualism*. 13 (4), 459–475.

Pyun, K-S. 2005. A model of interlanguage analysis: The case of Swedish by Korean speakers. In B. Hufeisen, & R. Fouser (Eds.), *Introductory Readings in L3* (pp. 55–70). Tübingen: Stauffenberg Verlag.

R Core Team. 2014. *R: A language and environment for statistical computing (Version 3.4.0) [Computer program]*, https://www.r-project.org.

Ringbom, H. 1987. *The Role of the Mother Tongue in Foreign Language Learning*. Clevedon: Multilingual Matters.

Ringbom, H., & Jarvis, S. 2009. The importance of cross-linguistic similarity in foreign language learning. In H. Michael, & J. Catherine (Eds.), *The Handbook of Language Teaching* (pp. 106–118). West Sussex: Wiley-Blackwell.

Shimizu, K. 1993. Heisa shiin no onseiteki tokutyou: yuseisei/museiseiteki gengokan hikaku nitsuite [A cross-language study of phonetic characteristics of stop consonants: With reference to voicing contrasts]. *Journal of Asian and African Studies* [*Azia/afurika gengo bunka kenkyu*]. 45, 163–176.

Tremblay, M. C. 2007. L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals. Paper presented at the Satellite Workshop of ICPhS XVI, Freiburg, Germany, 3–4 August 2007.

Williams, S., & Hammarberg, B. 1998. Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics*. 19, 295–333, DOI: 10.1093/applin/19.3.295.

Wrembel, M. 2014. VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics*. 5, 750–770, http://hdl.handle.net/10593/12094.

Wu, Zongji 吴宗济. 1988. 普通话发音不送气音/送气音区别的实验研究 [Aspirated vs. non-aspirated stops and affricates in Standard Chinese]. *中国语言学报 [Journal of Chinese Linguistics]*. 3, 256–283.

Wunder, E. 2010. Phonological cross-linguistic influence in third or additional language acquisition. *Proceedings of 6th New Sounds 2010*, Poznań, 566–571.

APPENDIX List of stimuli in the experiment

Language	monosyllable		disyllable	
Japanese	ぱ[pʰa]	ば[ba]	ぱさ[pʰasa]	ばさ[basa]
	た[tʰa]	だ[da]	たさ[tʰasa]	ださ[dasa]
	か[kʰa]	が[ga]	かさ[kʰasa]	がさ[gasa]
Russian	па[pa]	ба[ba]	пабо[pabo]	бабо[babo]
	та[ta]	да[da]	тадо[tado]	дадо[dado]
	ка[ka]	га[ga]	каго[kago]	гаго[gago]
Spanish	pa[pa]	ba[ba]	pamo[pamo]	bamo[bamo]
	ta[ta]	da[da]	tapu[tapu]	dapu[dapu]
	ca[ka]	ga[ga]	cami[kami]	gami[gami]

*The phonetic stimuli are in the word-initial syllables.