Article

second language research

Second Language Research
1–25

DOI: 10.1177/02676583221079147
journals.sagepub.com/home/slr
SAGE

Acquiring Italian stop consonants: A challenge for Mandarin Chinese-speaking learners

Qiang Feng and M. Grazia Busà
Università degli Studi di Padova, Padova, Italy

Abstract

The acquisition of Italian stop consonants by Mandarin Chinese-speaking learners has hardly been investigated. This study was designed to fill this gap. To investigate Chinese learners' acquisition patterns of Italian voiced and voiceless stops, a perception experiment and a production experiment were conducted. Twenty Mandarin Chinese-speaking undergraduate students majoring in Italian, five native Italian and five native Mandarin speakers served as participants in the perception experiment; and an equal number of participants with the same language backgrounds served as participants in the production experiment. In the perception experiment, the participants had to identify the stimuli in three continua (i.e. bilabial, alveolar and velar) where voice onset time (VOT) values ranged from −50 ms to 90 ms in 10 ms steps. In the production experiment, data were collected from a reading task in which the participants were asked to read the target words with word-initial stops in carrier-sentences; the VOT and closure durations were measured. The results show that, in perception, Chinese learners have difficulty differentiating between Italian voiced and voiceless stops; in production, Italian voiced rather than voiceless stops represent a challenge for Chinese learners. The results are in line with the predictions made by the Perceptual Assimilation Model-L2 (PAM-L2) and the Speech Learning Model (SLM), as well as with most other studies focusing on the acquisition of stops of 'true-voice languages' by Chinese learners.

Keywords

closure duration, Italian, L2 speech acquisition, Mandarin Chinese, stop consonants, voice onset time (VOT)

Corresponding author:
Qiang Feng, Università degli Studi di Padova, Via E. Vendramini 13, Padova, 35137, Italy
Email: qiang.feng@phd.unipd.it

I Introduction

The acquisition of stop consonants is a crucial issue in foreign language learning for Mandarin Chinese speakers. This has been investigated in various second language (L2) or third language (L3) learning contexts. However, to the best of our knowledge, the L2 Italian context has hardly been investigated. To fill this gap, the present study examines the perception and production of Italian word-initial stops by Mandarin Chinese-speaking learners (henceforth Chinese learners) as compared to native Italian and Mandarin speakers.

Both Standard Italian and Mandarin Chinese show a two-way laryngeal contrast between stops. However, the features of contrast are different. Specifically, in Standard Italian, stops (bilabial, alveolar and velar) are distinguished by voicing, and can be subdivided into voiced and voiceless (Kramer, 2009). On the other hand, in Mandarin Chinese, stops (bilabial, alveolar and velar) are voiceless, and are further subdivided into aspirated and unaspirated (Duanmu, 2007). Thus, following the categorization in Beckman, Jessen and Ringen (2013), Standard Italian and Mandarin Chinese belong respectively to 'true-voice' and 'aspirating' languages.

Chinese learners' acquisition of stop consonants in other true-voice languages, that is, French, Spanish (Nasukawa, 2005) and Russian (Beckman et al., 2013), has been the object of a number of investigations, which are reviewed in the following paragraphs.

In French, the average VOT (Lisker and Abramson, 1964) value of voiced stops is −95.7 ms; for voiceless stops, VOT values range between 17.9 ms and 62.2 ms depending on the place of articulation and vowel context (Nearey and Rochet, 1994). In the perception of French stops, Chinese learners locate their crossover boundaries between voiced and voiceless stops at higher VOT values than native French speakers (Rochet and Chen, 1992). Due to this, they tend to perceive both voiced and voiceless stops as voiced (Rochet and Chen, 1992), and therefore have difficulty distinguishing perceptually between them (Zhang, 2013). In production, Chinese learners have difficulty producing French voiced stops correctly, as these consonants tend to be replaced by Mandarin Chinese voiceless unaspirated stops and so are produced devoiced. Moreover, Chinese learners tend to produce French voiceless stops with smaller VOT values than native French speakers (Gabriel, Kupisch & Seoudy, 2016; Zhang, 2012).

In Spanish, mean VOT values range between −108 ms and −138 ms for word-initial voiced stops, and between 4 ms and 29 ms for voiceless stops (Lisker and Abramson, 1964). In the perception of Spanish stops, Chinese learners display higher crossover values than monolingual Spanish speakers (Liu and Cebrian, 2016). Thus, Chinese learners have more difficulty identifying correctly Spanish voiceless than voiced stops as both types of consonants tend to be categorized as voiced (Liu, Zeng and Lu, 2019). In production, Chinese learners tend to map Spanish voiced and voiceless stops to Mandarin Chinese unaspirated and aspirated stops respectively (Chen, 2007); alternatively, they produce Spanish voiced and voiceless bilabial stops respectively with negative and slightly positive VOT values (Liu and Cebrian, 2016).

In Russian, VOT values for voiced stops range from −70 ms to −78 ms, and for voiceless stops from 18 ms to 38 ms (Ringen and Kulikov, 2012). In perception, Chinese learners assimilate both Russian voiced and voiceless stops to Mandarin Chinese

voiceless unaspirated stops (Yang, Chen and Xiao, 2022), and therefore have difficulty perceiving the difference between Russian voiced and voiceless stops (Liu et al., 2019; Yang et al., 2022). In production, they acquire Russian voiceless but not voiced stops (Yang et al., 2022).

Similarly to French, Spanish and Russian, in Japanese the primary distinction between the two types of stops is voicing, with voiced stops showing negative VOT values ranging from −75 ms to −89 ms (Shimizu, 1996). However, Japanese voiceless stops present a slightly different scenario. That is, Japanese voiceless stops show positive VOT values ranging between 30 ms and 66 ms (Shimizu, 1996) or between 28.5 ms and 56.7 ms (Riney et al., 2007). Therefore, Japanese voiceless stops are described either as 'moderately aspirated' (Shimizu, 1996: 27) or as having 'an intermediate degree of aspiration' (Riney et al., 2007: 439). Possibly due to this, Chinese learners have an accuracy rate of about 70% in differentiating perceptually between Japanese voiced and voiceless stops (Hu, 2020). In production, Chinese learners tend to replace Japanese voiced and voiceless stops respectively with Mandarin Chinese voiceless unaspirated and aspirated stops (Jiang, 2020).

To sum up, the studies above suggest that in the perceptual differentiation between voiced and voiceless stops, Chinese learners seem to use the feature of aspiration instead of voicing. This leads them to confound voiced stops with their voiceless counterparts, especially when the latter are unaspirated. In production, voiced stops (which are absent in Mandarin Chinese phonology) appear to pose a greater challenge to Chinese learners than voiceless stops.

As for the acquisition of Italian stops by Chinese learners, to the best of our knowledge, this has been investigated in two studies. Xu (2019) examined Chinese learners' production of Italian word-initial bilabial and alveolar stops. Sun and Profita (2020) also investigated the production of Italian word-initial stops by Chinese learners in a study focused on the cross-linguistic influence of Chinese learners' L2 (English) on L3 (Italian) acquisition. Both studies show that Chinese learners tend to replace Italian voiced stops with Mandarin Chinese unaspirated stops, and acquire Italian voiceless stops well. However, both Xu (2019) and Sun and Profita (2020) are preliminary investigations, and have several limitations. First, both studies investigated only the production and not the perception data, which are essential for a pronunciation acquisition study. Second, in neither study were the results supported by statistical analyses. Third, both investigations used a relatively small number of Chinese learners, respectively eleven (Xu, 2019) and six (Sun and Profita, 2020), which might raise the concerns of Type S and Type M errors in light of the replication and reproducibility crisis.

In all the studies reviewed above, the conclusions were based mainly on the examination of VOT. Stop closure duration, a parameter that, alongside VOT, is closely related to the realization of stop consonants, has been the object of fewer investigations. To our knowledge, Ding et al. (2019) is the only investigation that has explicitly dealt with the closure durations in Chinese learners' production of English stops. This study shows that, in comparison to native English speakers, Chinese learners have longer closure durations for English unaspirated stops, and similar closure durations for English aspirated stops. However, since both English and Mandarin Chinese are aspirating languages, the closure duration patterns of Chinese learners' stop consonant production in true-voice languages remain an almost unresearched area.

Table 1. Voice onset time (VOT) values (in ms; SDs in parentheses if available) reported in the literature for Standard Italian and Mandarin Chinese stops.

Standard Italian	[b]	[d]	[g]	[p]	[t]	[k]
Vagges et al., 1978	−95	−50	−85	12	17	30
Bortolini et al., 1995	−73.7 (40)	−79.9 (38.8)	−66.9 (43.5)	11.3 (3.5)	19.3 (5.4)	34.1 (12.6)
Mandarin Chinese	[p]	[t]	[k]	[p^h]	[t^h]	[k^h]
Shimizu, 1996	7 (2.3)	12 (2.1)	19 (3.8)	96 (13.3)	98 (16.1)	112 (20.7)
Chao & Chen, 2008	14	16	27	82	81	92

The present study investigates Chinese learners' perception and production mechanisms in the acquisition of Italian stop consonants by examining both VOT and closure duration. In the next section, we look at the differences existing between Italian and Mandarin Chinese stops and formulate our hypotheses in light of these differences as well as current L2 speech acquisition theories.

II Stop consonants in Standard Italian and Mandarin Chinese

The phonological system of Standard Italian has three voiced and three corresponding voiceless stops; the voiced stops [b, d, g] are produced with vocal fold vibration during the closure, while the voiceless stops [p, t, k] are articulated without voicing and with short-lag VOT[1] (Kramer, 2009). In Mandarin Chinese, all stops are voiceless, with a distinction based on aspiration: [p, t, k] are voiceless unaspirated and [p^h, t^h, k^h] are voiceless aspirated (Duanmu, 2007). The VOT values reported in the literature for the stops in the two languages are shown in Table 1. Though the data are not always in agreement with each other, possibly due to different experimental designs, they do show that the two languages have clearly different stop categories.

According to Lisker and Abramson (1964) and Keating (1984), stops fall into three broad categories, namely 'voicing lead' (stops with negative VOT), 'short-lag' (stops with VOT of 0–35 ms) and 'long-lag' (stops with VOT larger than 60 ms). In light of this classification, Italian voiced stops fall into the 'voicing lead' category, and the voiceless ones into the 'short-lag' category; Mandarin Chinese voiceless unaspirated and aspirated stops can be respectively classified as 'short-lag' and 'long-lag' stops.

Since the primary distinction between the stops in Mandarin Chinese is aspiration rather than voicing, it is very possible that Chinese learners may apply the feature of aspiration to their perception of Italian stops, as they do with other true-voice languages. However, since both Italian voiced and voiceless stops are unaspirated, it is likely that to Chinese learners both categories of stops sound close to Mandarin Chinese voiceless unaspirated stops, leading to difficulty in distinguishing perceptually between Italian voiced and voiceless stops. Following the Perceptual Assimilation Model-L2 (PAM-L2; Best and Tyler, 2007), this situation could be interpreted as a case of 'Single Category (SC) assimilation', that is, two L2 sounds are assimilated to the same L1 sound category, leading to a poor discrimination performance.

Table 2. Standard Italian and Mandarin Chinese orthography for stop consonants.

Phonetic transcription		[b]	[d]	[g]	[p]	[t]	[k]	[p^h]	[t^h]	[k^h]
Orthography	Italian	<b>	<d>	<g>	<p>	<t>	<c>			
	Chinese				<b>	<d>	<g>	<p>	<t>	<k>

The perceptual assimilation of L2 sounds can be considered a case of 'equivalence classification', a process that, according to the Speech Learning Model (SLM; Flege, 1995, 1996, 2002), leads foreign language learners to approximate L1 and L2 sounds in production. In the case of Italian stops, it is plausible that Chinese learners may ignore perceptually the difference between Italian voiced [b, d, g] and voiceless [p, t, k], and produce both categories of stops similarly to Mandarin Chinese voiceless unaspirated [p, t, k], as it happens in L2 French (Gabriel et al., 2016; Zhang, 2012) and Russian (Yang et al., 2022).

In addition to the perceptual difficulty, the different orthography for stops in the two languages may be a source of confusion for Chinese learners. In Standard Italian, <b, p>, <d, t> and <g, c> are used for writing [b, p], [d, t] and [g, k]. In Mandarin Chinese, the orthographic forms[2] <b, p>, <d, t> and <g, k> are used for writing [p, p^h], [t, t^h] and [k, k^h] respectively (for a comparison, see Table 2). The confusion caused by the different orthographic conventions may add to the perceptual difficulty, and cause Chinese learners, trying to differentiate Italian voiced [b, d, g] from voiceless [p, t, k], to transfer the feature of contrast for Mandarin Chinese stops (i.e. aspiration) to Italian stops. Thus, it can be hypothesized that, similarly to what happens with L2 Japanese (Jiang, 2020) and Spanish (Chen, 2007), Chinese learners may produce Italian voiced [b, d, g] like Mandarin Chinese unaspirated [p, t, k], and Italian voiceless [p, t, k] like Mandarin Chinese aspirated [p^h, t^h, k^h].

As for closure duration, three factors may play a role in determining Chinese learners' production of Italian stops. Firstly, in Italian voiced stops tend to have shorter closure durations than voiceless stops (Coretta, 2019; Esposito, 2002); in Mandarin Chinese voiceless unaspirated stops tend to have longer closure durations than aspirated stops (Svantesson, 1987). We hypothesize that Chinese learners either produce both Italian voiced and voiceless stops like Mandarin Chinese unaspirated stops, or produce the two categories similarly to Mandarin Chinese unaspirated and aspirated stops respectively. Thus, correspondingly, we hypothesize that Chinese learners either produce Italian voiced and voiceless stops with similar closure durations, or produce longer closure durations for Italian voiced than for voiceless stops.

Secondly, bilabial stops tend to have longer closure durations than alveolar and velar stops (Cho and Ladefoged, 1999). Since this tendency holds true for both Italian (Esposito, 2002) and Mandarin Chinese stops (Svantesson, 1987), we hypothesize that in producing stops both native Italian speakers and Chinese learners follow the common durational pattern.

Thirdly, the closure durations of Italian stop consonants have been found to be inversely related to speaking rates (Pickett, Blumstein and Burton, 1999). That is, the

slower one speaks, the longer stop closure durations are. Since foreign language learners usually speak slower than native speakers, our hypothesis is that in the production of Italian stops Chinese learners tend to produce longer stop closure durations than native Italian speakers.

To sum up, we hypothesize that: in perception, Chinese learners have difficulty differentiating between Italian voiced and voiceless stops (H1); in production, in terms of both VOT and closure duration, Chinese learners either produce both Italian voiced and voiceless stops similarly to the corresponding Mandarin Chinese unaspirated ones (H2), or produce Italian voiced and voiceless stops respectively like Mandarin Chinese unaspirated and aspirated stops (H3). Concerning solely stop closure duration, we expect that all participants follow the same durational pattern in producing stops, and Chinese learners produce longer closure durations than native Italian speakers due to the former having slower speaking rates than the latter (H4).

To investigate these hypotheses, a perception experiment and a production experiment were run. The following sections describe the methods and results of each experiment.

III Perception experiment

1 Method

a Participants. Three groups of participants were involved in the perception experiment. None of them reported any hearing impairment at the time of the experiment.

The experimental group (EXP) consisted of 20 Chinese students (Female = 18, Male = 2, Mean age = 20.4, Age range = 20–21). They were all third-year undergraduate students majoring in Italian at Dalian University of Foreign Languages in China. Their gender distribution reflects the imbalance of the students' enrollment in the degree course. Their Italian proficiency could be approximated to the B1 level of the Common European Framework of Reference for Languages (CEFR).[3]

The first control group (IT) consisted of five monolingual native Italian-speaking high school/undergraduate students (Female = 2, Male = 3, Mean age = 19.4, Age range = 19–20) from the Veneto region (North-East Italy). To our knowledge, with regard to the VOT patterns of word-initial singleton stops, regional Italian from Veneto has not been reported to diverge from Standard Italian.

The second control group (MC) consisted of five monolingual native Mandarin Chinese-speaking undergraduate students (Female = 3, Male = 2, Mean age = 20.8, Age range = 20–21) from the northern dialect area of China where the dialectal influence is minimal in terms of Mandarin Chinese stops (Yuan, 2001).

b Materials. To test the three groups' differences in category boundaries for stop consonants, the VOT continua of bilabial, alveolar and velar stops were prepared.

In the first place, we determined the VOT ranges of the continua. Brady and Darwin (1978) and Keating, Mikoś and Ganong III (1981) examined a series of alveolar continua with different VOT ranges, and found that the perceptual boundaries of the same listeners were substantially shifted due to 'range effects'. That is, even if two groups

of listeners share identical perceptual patterns, their category boundaries for two VOT continua of the same place of articulation might be significantly different if the two continua have different VOT ranges. Based on this, to eliminate a possible bias caused by different VOT ranges (i.e. 'range effects'), we decided to use an identical VOT range for all the VOT continua and all the participants involved in the present perception experiment. In this way, we expected to be able to ascertain that the within- and/or between-group differences in perceptual boundaries (if any) were not caused by different VOT ranges, but by different perceptual patterns.

The VOT range of the continua was initially set at −60 ms to 90 ms. Changes in VOT were implemented in 10 ms steps. The decision to use these values was made following Flege and Eefting (1986, 1987a, 1987b, 1988). In each of these studies, the authors investigated a true-voice language vs. an aspirating one (i.e. Spanish vs. English, Dutch vs. English). As claimed in Flege and Eefting (1987b: 72), the range of −60 ms to 90 ms with 10 ms steps 'provided exemplars of the three modal VOT categories used to implement stops', namely voicing lead, short-lag and long-lag. Therefore, it seemed appropriate to use the same range of VOT for the three stop categories in Italian and Mandarin Chinese.

The VOT continua were generated following the tutorial in Winn (2020). First, the original sounds were recorded. A native Mandarin Chinese speaker from the northern dialect area of China was recruited. He was instructed to produce three pairs of monosyllables, namely [pa] vs. [p^{h}a], [ta] vs. [t^{h}a], and [ka] vs. [k^{h}a] with equivalent perceived loudness. Each syllable was produced in isolation and saved as a separate audio file. The recordings were collected in a quiet setting using a Zoom H4n Pro voice recorder with a sampling rate of 44.1 kHz and 16-bit resolution.

Subsequently, each pair of the original sounds was manipulated using the script provided by Winn (2020) in Praat (Boersma and Weenink, 2020). In the startup window of the script, we set the minimum and maximum VOT values respectively at −60 ms and 90 ms, and the number of VOT steps at 16. The other parameters were left as default. Then we initialed the generation procedure. After the timing landmarks (i.e. the start of the burst and the end of the aspiration in the aspirated stops, and the vowel onset in the unaspirated stops) were manually selected from the original sounds, the script automatically generated three continua that increased in 10 ms steps from −60 ms to 90 ms ranging respectively from [ba] to [p^{h}a], [da] to [t^{h}a], and [ga] to [k^{h}a].

The last step was the assessment of the continua. When checking the generated audio files, we found that the tokens with VOT values of −60 ms sounded somewhat unnatural. For this reason, they were excluded from the continua. In this way, in each continuum there were 15 tokens. Their VOT values ranged from −50 ms to 90 ms in 10 ms steps. In total, 45 different tokens (3 places of articulation × 15 tokens = 45) were created.

c Procedure. An Italian version (for the EXP and IT groups) and a Mandarin Chinese version (for the MC group) of the identification tests were developed using Experiment-MFC 4 in Praat. In both versions of the tests, each of the 45 tokens was repeated three times. The 135 stimuli (45 tokens × 3 repetitions = 135) were randomly presented to the

participants with a 1s interstimulus interval. After every 30 stimuli there was a break. At any time during the break, the participants could click, when ready, on the *Continua*/<继续> 'continue' button to enter the next block of stimuli. The tests presented six option buttons on the computer screen. In the Italian version, the six option buttons were <ba, pa, da, ta, ga, ca>. In the Mandarin Chinese version, the six option buttons were presented in original Chinese characters followed by their transliteration in Pinyin, namely <巴> *ba*, <趴> *pa*, <搭> *da*, <他> *ta*, <嘎> *ga*, <咖> *ka*. The participants were asked to click on the option button that corresponded to the stimulus heard. Though they were told to make their choices without thinking too much, the participants were allowed to listen to each stimulus for a maximum of three times.

Due to the lockdowns and movement restrictions caused by the Covid-19 pandemic, we were forced to run the perception experiment remotely. First, we sent all the materials to the participants and asked them to download them in advance. Then we invited the participants, divided into two groups by their native language, to attend an online meeting in which we instructed them how to install Praat on their computers, how to start the experiment with Praat, and how to extract the final results at the end of the test. All the participants were asked to take the test in a quiet environment with their computer headphones on.

In order to activate the Italian mode (Grosjean, 2007) of the EXP group, the participants' perception experiment was conducted immediately after an online lesson in Italian given by a native Italian teacher. For the EXP group and the IT group, all the instructions in the identification test were given in Italian. For the MC group, the instructions were given in Mandarin Chinese. The participants were asked to send us their final results once they were finished with the experiment.

To check the reliability of the final results, we looked at the participants' misperceptions of the stimuli. By misperception we refer to the cases when a stimulus in one of the VOT continua was perceived as a stop of another continuum (e.g. a stimulus in the bilabial continuum was perceived as an alveolar or velar stop). In total, we found 238 misperceptions (57 for the bilabial continuum, 170 for the alveolar continuum, 11 for the velar continuum) out of 4,050 responses (135 stimuli $\times$ [20+5+5] participants = 4,050). This low percentage of misperceptions (238/4,050 = 5.9%) showed that overall the participants carried out the identification tests in the appropriate way.

d Analyses. Following Caramazza et al. (1973: 424), we define a category boundary as 'the crossover point marking the VOT value at which 50% of the responses are for one phoneme category and 50% for the other'. In order to determine the three groups' category boundaries for each VOT continuum, the valid responses obtained in the identification tests were first sorted out into three blocks by continuum (i.e. bilabial, alveolar and velar). The responses for each continuum were then further divided into three sub-blocks by group (i.e. EXP, IT and MC). Finally, for each of the nine sub-blocks (3 continua $\times$ 3 groups = 9), the <ba/da/ga> response percentages at every single stimulus were calculated.

The misperceived responses were excluded from the statistical analyses. The remaining valid responses were recoded into a binary variable: the <ba, da, ga> were coded as

'1' and the <pa, ta, ca/ka> as '0'. Here and in what follows we use the forms <ca/ka> and <c/k> because of the different orthographic conventions used in Standard Italian and Mandarin Chinese (see Table 2); note that Italian <c> corresponds to short-lag [k] and Mandarin Chinese <k> corresponds to long-lag [k^h]. A generalized linear mixed-effects model (GLMM) with a binomial link function was then applied to the responses using the lme4 package 1.1.26 (Bates et al., 2015) in R 3.6.3 (R Core Team, 2020). For this GLMM, the fixed factors were Group (three levels: EXP, IT, and MC), Continuum (three levels: bilabial, alveolar, and velar), VOT (treated as continuous and centered at zero), and their interactions. The random intercept was Participant. The main effects of the fixed factors were assessed by the Type II Wald chi-squared tests using the car package 3.0.10 (Fox and Weisberg, 2019). Post-hoc comparisons of contrasts were performed using the emmeans package 1.5.3 (Lenth, 2020).

2 Results

In total, we had 3,812 valid responses (4,050 responses − 238 misperceptions = 3,812). Based on these data, the EXP, IT and MC groups' average labeling functions for the bilabial, alveolar and velar VOT continua are plotted in Figure 1. As the figure shows, for the bilabial continuum the VOT values that were closest to the 50% crossover points (henceforth near-crossover values) were 0 ms for the IT group, and 20 ms for the EXP and MC groups. For the alveolar continuum the near-crossover values were 20 ms for the IT group, and 30 ms for the EXP and MC groups. For the velar continuum these values were 30 ms for the IT group, and 40 ms for the EXP and MC groups.

The GLMM applied to the three groups' responses yielded main effects on Group ($\chi^2(2) = 13.31$, $p = 0.001$), Continuum ($\chi^2(2) = 80.38$, $p < 0.001$) and VOT ($\chi^2(1) = 496.92$, $p < 0.001$). There were significant interactions between Group and Continuum ($\chi^2(4) = 23.10$, $p < 0.001$), between Continuum and VOT ($\chi^2(2) = 42.72$, $p < 0.001$), between Group and VOT ($\chi^2(2) = 17.34$, $p < 0.001$), and between Group, Continuum and VOT ($\chi^2(4) = 15.94$, $p = 0.003$).

To see if any differences existed between the EXP group and the other two control groups in terms of their category boundaries, post-hoc comparisons of contrasts were implemented at all the near-crossover values of the EXP group. For clarity, the results are summarized in Table 3. As the results show, for all the three continua, no significant differences were found between the EXP group and the MC group. This indicates that these two groups shared the same category boundaries for all the three continua. Moreover, the EXP and MC groups were always significantly different from the IT group. This suggests that, for all the three continua, the EXP and MC groups' crossover values were significantly different from those of the IT group.

3 Discussion

The overlap between the Chinese learners' and the native Mandarin speakers' perceptual category boundaries for all the three VOT continua suggests that Chinese learners'

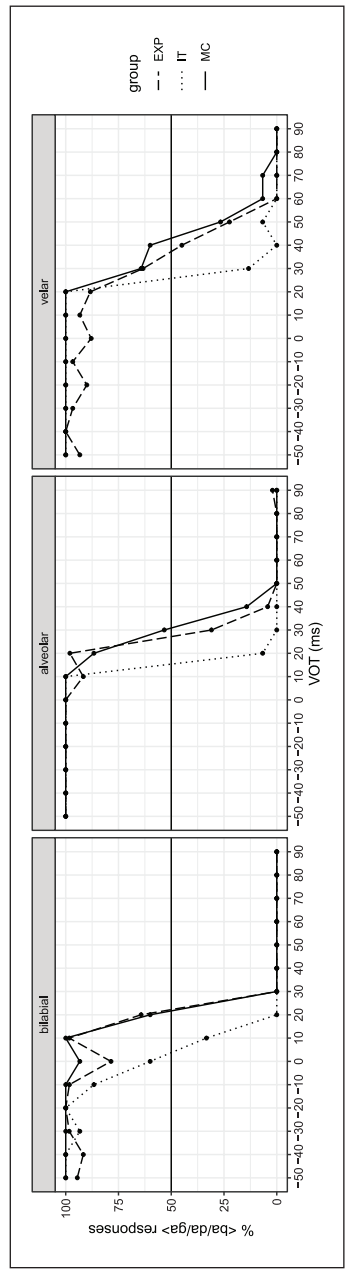

Figure 1. The EXP, IT and MC groups' average labeling functions for the bilabial, alveolar and velar VOT continua.
Notes. VOT = voice onset time. EXP = experimental group. IT = first control group. MC = second control group.

Table 3. Summary of the results of the comparisons of contrasts at the EXP group's near-crossover values for the bilabial, alveolar and velar continua.

Continuum	VOT (ms)	Group	Estimate	*SE*	*z* ratio	*p* value
Bilabial	20	EXP vs. IT	2.68	0.73	3.68	0.0073*
		EXP vs. MC	−0.41	0.59	−0.69	0.9989
		IT vs. MC	−3.09	0.89	−3.47	0.0154*
Alveolar	30	EXP vs. IT	22.27	6.68	3.33	0.0242*
		EXP vs. MC	−0.54	0.58	−0.93	0.9916
		IT vs. MC	−22.81	6.70	−3.41	0.0190*
Velar	40	EXP vs. IT	3.94	1.23	3.20	0.0367*
		EXP vs. MC	−0.78	0.49	−1.58	0.8145
		IT vs. MC	−4.71	1.29	−3.64	0.0083*

Notes. * $p < 0.05$. VOT = voice onset time. EXP = experimental group. IT = first control group. MC = second control group.

perception of stop consonants is greatly affected by their L1 rules, such that they follow different ways to categorize stops as compared to native Italian listeners.

Regarding the exact crossover value, as suggested in Keating et al. (1981), it is supposed to be 0 ms VOT for speakers of true-voice languages. However, in the present perception experiment, the native Italian speakers' crossover values were always higher than 0 ms (they were between 0 ms and 30 ms). Why is that? The most plausible cause is Italian speakers' sensitivity to 'range effects'. As claimed in Keating et al. (1981), in perceptual identification tests, for a VOT continuum with appreciable numbers of voiceless stimuli and few prevoiced stimuli (as in the present study), the category boundaries of listeners of true-voice languages will diverge from their actual crossover value, namely 0 ms VOT, and shift towards short-lag VOT values.

As for the Chinese learners and the native Mandarin speakers, their crossover VOT values varied between about 20 ms and 40 ms as a function of the places of articulation of the VOT continua. These crossover values are considered reliable for two reasons. First, in comparison to speakers of true-voice languages, those of aspirating languages are much less prone to 'range effects'. That is, their category boundaries are quite stable and almost unaffected by the VOT ranges of acoustic continua (Keating et al., 1981). Second, this result is generally consistent with other studies focusing on the categorical perception of Mandarin Chinese stops (e.g. Rochet and Yanmei, 1991; Yang and Fang, 1984; Zhang, 2014). Therefore, we can say that Chinese learners have higher crossover values (about 20–40 ms) than native Italian speakers (0 ms).

Since the VOT values of Italian stops are generally smaller than the crossover values of Chinese learners (see Table 1), it is conceivable that, when perceiving Italian stops, Chinese learners tend to categorize both voiced and voiceless stops within the same category, namely the unaspirated one. This finding is compatible with our H1 formulated according to the PAM-L2 theory (Best and Tyler, 2007); that is, Chinese learners have difficulty differentiating between Italian voiced and voiceless stops.

IV Production experiment

1 Method

a Participants. The production experiment, like the perception experiment, involved three groups of participants. They were highly similar to the participants in the perception experiment in terms of language background. Specifically, the experimental group (EXP) consisted of 20 Chinese third-year undergraduate students (Female = 17, Male = 3, Mean age = 20.5, Range = 20–21) majoring in Italian at Dalian University of Foreign Languages in China. The first control group (IT) consisted of five monolingual native Italian undergraduate students (Female = 4, Male = 1, Mean age = 20.2, Age range = 20–21) from the Veneto region in the North-East of Italy. The second control group (MC) consisted of five monolingual native Mandarin Chinese undergraduate students (Female = 4, Male = 1, Mean age = 20.0, Age range = 19–21) from the northern dialect area of China. The participants reported no speech impairment at the time of the experiment. None of the participants participated in the perception experiment.

b Materials. An Italian version and a Mandarin Chinese version of the stimuli were prepared. In the Italian version two frequently used Italian words were selected as target stimuli for each of the six Italian stops (see Appendix 1). All these 12 words (6 stops × 2 words = 12) were disyllables with stress on the first syllable; the stops occurred in word-initial position and were followed by [a]. The 12 target words consisted of four minimal pairs and two quasi-minimal pairs contrasting in stop types. To prevent the participants from grasping the experiment purpose, 16 other disyllabic words were used as distractors (see Appendix 1). The distractors consisted of four minimal pairs contrasting in consonant length and four minimal pairs contrasting in [r–l]. So, there was a total of 28 word stimuli (6 stops × 2 target words + 16 distractors = 28) for the Italian version. All the word stimuli were first inserted in the carrier phrase *Leggo* ___ *bene* 'I read ___ well', repeated twice in a randomized order, and finally printed on a paper sheet.

The Mandarin Chinese target stimuli were also 12 frequently used disyllabic words with stops in word-initial position followed by [a]. Moreover, all the first syllables were of the first tone (see Appendix 1). Since there were no minimal pairs contrasting in stop types among the Mandarin Chinese target stimuli, we deemed unlikely that the participants would easily grasp the experiment purpose. For this reason, the 16 Mandarin Chinese disyllabic distractors were selected without any specific criterion (see Appendix 1). The 28 Mandarin Chinese word stimuli (6 stops × 2 target words + 16 distractors = 28) were presented in original Chinese characters. They were embedded in the carrier sentence <我说___这个词> 'I say ___ this word', repeated twice randomly, and printed on a sheet of paper.

c Procedure. The sheet with the Italian version of the stimuli was given to the EXP group and the IT group to read. The other sheet with the Mandarin Chinese stimuli was given to the MC group. All the participants were instructed to read the sentences in a natural way at a normal speed. They were also asked to make a short break after every 10 sentences. The recording of the IT group took place in the Language and Communication

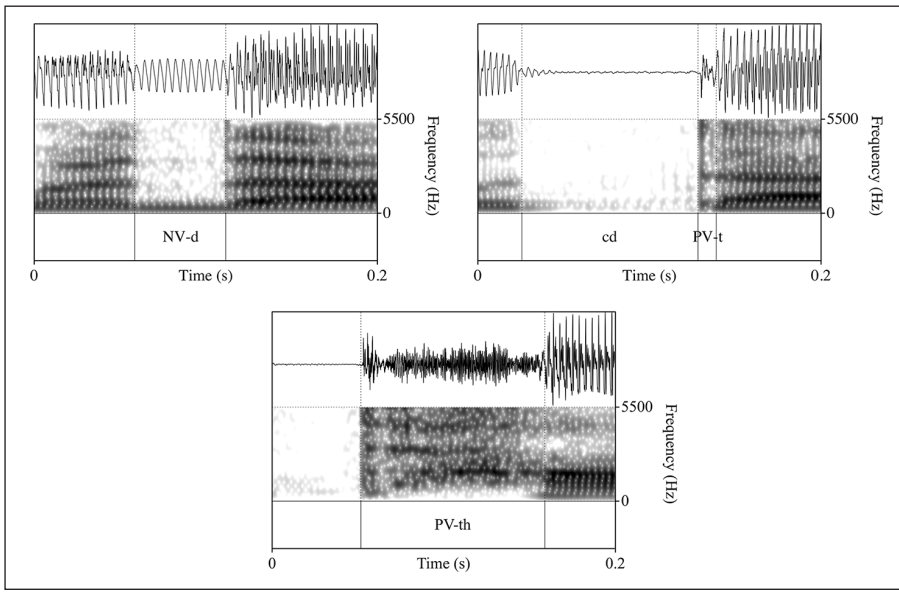

Figure 2. Acoustic waveforms and spectrograms at a 200 ms time scale of (i) the VOT of an Italian voiced [d] produced by the participant IT-2 in the IT group (upper left panel); (ii) the closure duration and VOT of an Italian voiceless unaspirated [t] produced by the participant EXP-20 in the EXP group (upper right panel); and (iii) the VOT of a Mandarin Chinese aspirated [t^h] produced by the participant MC-5 in the MC group (bottom panel).
Notes. VOT = voice onset time. EXP = experimental group. IT = first control group. MC = second control group. NV = negative VOT. PV = positive VOT. cd = closure duration.

Lab of the University of Padova in Italy, using a Roland R09 voice recorder with a sampling rate of 44.1 kHz and 16-bit resolution. The recordings of the EXP group and the MC group were administered in a quiet setting at Dalian University of Foreign Languages in China, using a Zoom H4n Pro voice recorder with a sampling rate of 44.1 kHz and 16-bit resolution.

d Annotation and measurement. The annotation and measurement of VOT were performed in Praat by examining the acoustic waveforms of the tokens elicited in the reading task, following Francis, Ciocca and Ching Yu (2003). The wave oscillation before the release burst of the stop was annotated as negative VOT for the voicing lead stops. The temporal span between the release burst and the following onset of periodicity in the waveform was labeled as positive VOT for the short-lag and long-lag stops. Here, the onset of periodicity was identified as 'the time of the zero-crossing preceding the upward-going portion of the first cycle of oscillation visible in the acoustic waveform' (Francis et al., 2003: 1027). In addition, we also annotated the closure durations of the stop consonants produced in the production experiment. The closure duration was identified as the time interval between the offset of the periodic wave of the preceding vowel and the release burst of the stop consonant. Sample graphs in Figure 2 show how the negative

Table 4. Mean, average, median VOT values (in ms; SDs in parentheses) and percentages of prevoicing of the <b, d, g> and <p, t, c/k> produced by the EXP, IT and MC groups.

	EXP	IT	MC
Category <b, d, g>:			
<b>	0.7 (59.7)	−74.9 (17.9)	16.2 (3.4)
<d>	6.4 (46.1)	−58.3 (13.3)	17.2 (5.3)
<g>	21.6 (30.4)	−61.0 (14.3)	27.4 (5.4)
Average	9.5 (47.6)	−65.1 (16.9)	20.3 (6.9)
Median	18.3	−63.8	18.2
% prevoicing	7.6%	100%	0%
Category <p, t, c/k>:			
<p>	16.9 (7.5)	15.1 (3.6)	104.9 (19.5)
<t>	21.0 (21.1)	20.3 (6.4)	104.4 (17.3)
< c/k >	26.4 (7.5)	41.1 (8.5)	103.1 (15.8)
Average	21.5 (14.0)	25.8 (13.1)	104.1 (17.3)
Median	18.8	20.4	103.1
% prevoicing	0.4%	0%	0%

Notes. VOT = voice onset time. EXP = experimental group. IT = first control group. MC = second control group.

VOT, positive VOT and closure duration were annotated. Finally, the VOT values and the closure durations were extracted with a Praat script (Lennes, 2002).

e *Analyses.* To facilitate the verification of our hypotheses regarding the three groups' production of VOT values and the production of closure durations, the stop consonants of Italian and Mandarin Chinese were divided into two categories based on their orthographic forms (see Table 2). That is, <b, d, g> were gathered as one category, and <p, t, c/k> as another. Two linear mixed models (LMMs) were applied respectively to the VOT values and the closure durations using the lme4 package 1.1.26 (Bates et al., 2015) in R 3.6.3 (R Core Team, 2020), with Group (three levels: EXP, IT, MC), Category (two levels: <b, d, g> and <p, t, c/k>), and their interaction as fixed factors, and Participant and Stimulus as random intercepts. The assessments of the main effects of the fixed factors were performed with the Type II Wald chi-squared tests using the car package 3.0.10 (Fox and Weisberg, 2019). Post-hoc Bonferroni pairwise comparisons were conducted using the emmeans package 1.5.3 (Lenth, 2020).

2 Results

a VOT. A total of 720 target tokens (6 stops × 2 target word stimuli × 2 repetitions × [20+5+5] participants = 720) were elicited. For the statistical analyses, 20 unmeasurable tokens were discarded, leaving us 700 valid tokens (720 target tokens − 20 unmeasurable tokens = 700).

As shown in Table 4, the IT group produced Italian word-initial <b, d, g> ([b, d, g]) and <p, t, c> ([p, t, k]) respectively as fully voiced and voiceless stops. The MC group

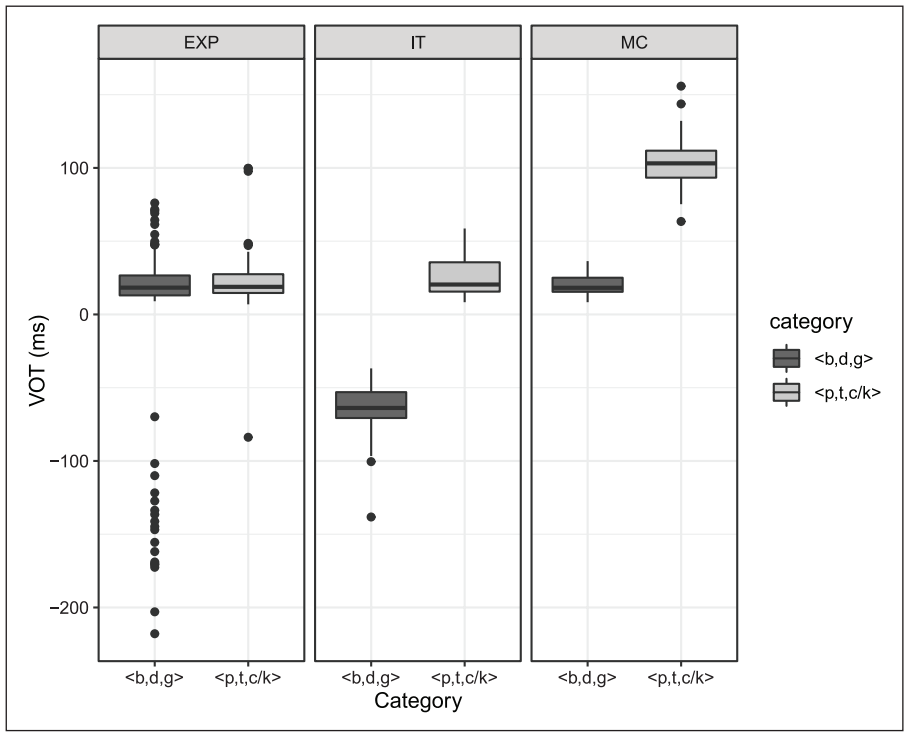

Figure 3. VOT distributions of the stops produced by the EXP, IT and MC groups.
Notes. VOT = voice onset time. EXP = experimental group. IT = first control group. MC = second control group.

produced Mandarin Chinese word-initial <b, d, g> ([p, t, k]) and <p, t, k> ([p^h, t^h, k^h]) respectively as unaspirated and aspirated stops. As for the EXP group, they produced Italian voiceless <p, t, c> ([p, t, k]) with short-lag VOT values. However, their production of Italian voiced <b, d, g> ([b, d, g]) was rather unstable, as can be seen from the large standard deviations. Specifically, they produced a small portion (7.6%) of Italian voiced stops with negative VOT values and a large portion (92.4%) with short-lag values (mean: 22.5 ms).

A closer inspection of the 18 Italian voiced stops produced with negative VOT values by the EXP group revealed that [b] and [d] were realized as voiced more often than [g]: out of 18 occurrences [b] was produced as voiced ten times, [d] six, and [g] two. Moreover, the 18 voiced stops were mainly realized by the participants EXP-16 (seven instances out of 18) and EXP-17 (eight instances out of 18) who were both from the northern dialect area of China. All the 18 stops were produced with rather long prevoicing (mean: −147.5 ms).

The VOT distributions in Figure 3 show that there were no overlaps either between the Italian voiced and voiceless stops produced by the IT group, or between the Mandarin Chinese unaspirated and aspirated stops produced by the MC group. However, the EXP group's Italian voiced and voiceless stops had a narrow distribution within the short-lag range.

For the statistical analyses, the VOT values were first normalized using the bestNormalize package 1.7.0 (Peterson and Cavanaugh, 2020). After fitting the LMM, the visual diagnostics of the histogram and the plot of residuals revealed no drastic violations of the assumptions of normality and homoscedasticity. The LMM yielded significant main effects on Group ($\chi^2(2) = 36.89$, $p < 0.001$), Category ($\chi^2(1) = 53.28$, < 0.001), and their interaction ($\chi^2(2) = 201.88$, $p < 0.001$).

Regarding the between-group differences in VOT, for <b, d, g>, pairwise comparisons showed that the EXP group was significantly different from the IT group ($\beta = 1.36 \pm 0.18$ *SE*, $t(34.4) = 7.58$, $p < 0.001$), but similar to the MC group ($\beta = -0.15 \pm 0.26$ *SE*, $t(38.6) = -0.56$, $p = 1.000$). The results show that the Chinese learners produced Italian voiced [b, d, g] like Mandarin Chinese unaspirated [p, t, k] and failed to produce Italian voiced stops. For <p, t, c/k>, pairwise comparisons showed that the EXP group was significantly different from the MC group ($\beta = -1.84 \pm 0.27$ *SE*, $t(38.8) = -6.96$, $p < 0.001$), but similar to the IT group ($\beta = -0.23 \pm 0.18$ *SE*, $t(34.0) = -1.30$, $p = 1.000$). The results show that the Chinese learners did not aspirate Italian voiceless [p, t, k], and produced them like the native Italian speakers.

Concerning the within-group differences in VOT, significant differences were found between the Italian <b, d, g> and <p, t, c> produced by the IT group ($\beta = -1.75 \pm 0.22$ *SE*, $t(25.2) = -7.87$, $p < 0.001$), and between the Mandarin Chinese <b, d, g> and <p, t, k> produced by the MC group ($\beta = -1.84 \pm 0.22$ *SE*, $t(24.9) = -8.34$, $p < 0.001$). However, no significant differences were found between the EXP group's Italian <b, d, g> and <p, t, c> ($\beta = -0.15 \pm 0.20$ *SE*, $t(17.2) = -0.73$, $p = 1.000$). Moreover, the EXP group produced both categories similarly to the MC group's <b, d, g> (EXP <b, d, g> vs. MC <b, d, g>: $\beta = -0.15 \pm 0.26$ *SE*, $t(38.6) = -0.56$, $p = 1.000$; EXP <p, t, c> vs. MC <b, d, g>: $\beta = -0.0002 \pm 0.26$ *SE*, $t(38.6) = -0.001$, $p = 1.000$). The results show that the Chinese learners did not produce Italian voiced and voiceless stops distinctively. They produced both categories similarly to Mandarin Chinese unaspirated stops.

b Closure duration. In total, we had 720 target tokens (6 stops × 2 target word stimuli × 2 repetitions × [20+5+5] participants = 720). For the statistical analyses, 13 unmeasurable tokens were discarded. Besides, 25 tokens produced with conspicuous hesitation were also discarded because during the intervals of hesitation it was impossible to know when the closures of the word-initial voiceless stops (17 occurrences), or of the voiced stops mispronounced as voiceless stops (8 occurrences) started. In this way, 682 effective tokens (720 target tokens − 38 discarded tokens = 682) were left for the statistical analyses.

As can be seen from Table 5, all of the three groups followed the common pattern for stop closure duration: their bilabial stops had longer closure durations than their alveolar and velar stops. Regarding the average closure durations, Figure 4 shows that the IT group produced shorter closure durations for Italian voiced <b, d, g> ([b, d, g]) than for voiceless <p, t, c> ([p, t, k]); the MC group had longer closure durations for Mandarin Chinese voiceless unaspirated <b, d, g> ([p, t, k]) than for voiceless aspirated <p, t, k> ([p^h, t^h, k^h]). As for the EXP group, their closure durations of Italian voiced stops were slightly longer than those of the voiceless ones. Moreover, the average closure durations of the EXP group were always longer than those of the IT and MC groups.

Table 5. Mean and average closure durations (in ms; SDs in parentheses) of the <b, d, g> and <p, t, c/k> produced by the EXP, IT and MC groups.

	EXP	IT	MC
Category <b, d, g>:			
<b>	122.3 (39.4)	74.9 (17.9)	97.0 (27.5)
<d>	116.4 (43.5)	58.3 (13.3)	84.2 (16.3)
<g>	107.9 (37.0)	61.0 (14.3)	75.7 (10.2)
Average	115.4 (40.3)	65.1 (16.9)	85.8 (21.2)
Category <p, t, c/k>:			
<p>	115.7 (42.4)	94.9 (20.3)	71.7 (14.1)
<t>	103.2 (32.8)	86.9 (14.5)	62.2 (19.7)
< c/k >	95.9 (40.6)	67.3 (11.2)	58.6 (14.3)
Average	104.9 (39.5)	82.8 (19.4)	64.3 (17.0)

Notes. EXP = experimental group. IT = first control group. MC = second control group.

For the statistical analyses, the closure duration values were first normalized using the best Normalize package 1.7.0 (Peterson and Cavanaugh, 2020). After fitting the LMM, the visual inspection of the histogram and the plot of residuals revealed no drastic deviations from the assumptions of normality and homoscedasticity. The LMM yielded significant main effects on Group ($\chi^2(2) = 23.37$, $p < 0.001$), Category ($\chi^2(1) = 11.11$, $p < 0.001$), and their interaction ($\chi^2(2) = 54.57$, $p < 0.001$).

Regarding the within-group differences in closure duration, pairwise comparisons showed significant differences between the Italian <b, d, g> and <p, t, c> produced by the IT group ($\beta = -0.68 \pm 0.18$ *SE*, $t(40.6) = -3.84$, $p = 0.006$), and between the Mandarin Chinese <b, d, g> and <p, t, k> produced by the MC group ($\beta = 0.82 \pm 0.18$ *SE*, $t(38.7) = 4.66$, $p < 0.001$). However, no significant differences were found between the EXP group's Italian <b, d, g> and <p, t, c> ($\beta = 0.28 \pm 0.14$ *SE*, $t(14.7) = 1.99$, $p = 0.978$). These results show that, in terms of closure duration, the native Italian and Mandarin speakers produced the two stop categories in their respective native languages distinctively. On the contrary, the Chinese learners confounded Italian voiced stops with the voiceless ones.

Regarding the between-group differences, for <b, d, g>, pairwise comparisons showed that the EXP group was significantly different from the IT group ($\beta = 1.48 \pm 0.27$ *SE*, $t(31.5) = 5.53$, $p < 0.001$); while for <p, t, c>, the EXP group was similar to the IT group ($\beta = 0.52 \pm 0.27$ *SE*, $t(31.6) = 1.93$, $p = 0.939$). Moreover, no significant differences were found either between the EXP group's <b, d, g> and the MC groups' <b, d, g> ($\beta = 0.69 \pm 0.29$ *SE*, $t(39.7) = 2.36$, $p = 0.354$), or between the EXP groups' <p, t, c> and the MC groups' <b, d, g> ($\beta = -0.41 \pm 0.29$ *SE*, $t(39.7) = -1.41$, $p = 1.000$). These results show that, in terms of closure duration, the Chinese learners approximated the L2 native norms in producing Italian voiceless but not voiced stops. Moreover, they produced both Italian voiced and voiceless stops similarly to Mandarin Chinese voiceless unaspirated stops.

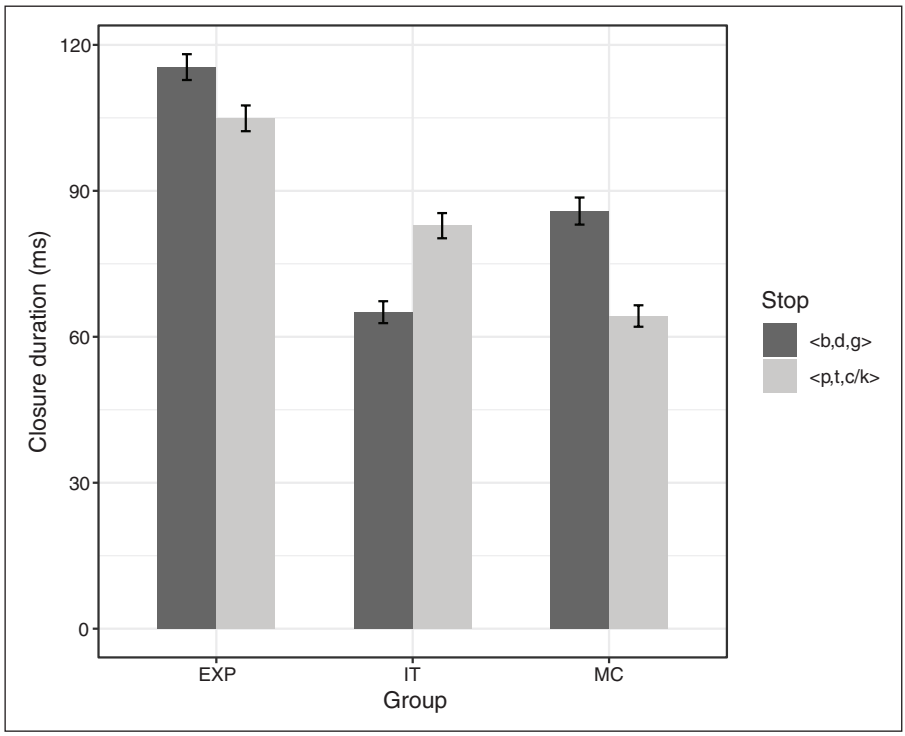

Figure 4. Average closure durations of the stops produced by the EXP, IT and MC groups.
Notes. EXP = experimental group. IT = first control group. MC = second control group.

3 Discussion

The results of the three groups' productions of VOT values and closure durations can be interpreted together. First of all, the productions of the native Italian and Mandarin speakers align with what is reported in the literature. That is, the stop consonants in both Italian and Mandarin Chinese can be divided into two distinctive categories. In Italian, the two categories are voiced and voiceless with voiced stops having negative VOT and relatively shorter closure durations, and voiceless stops having short-lag VOT accompanied by relatively longer closure durations. In Mandarin Chinese, the two categories are voiceless unaspirated and aspirated with the former having short-lag VOT accompanied by relatively longer closure durations, and the latter having long-lag VOT and relatively shorter closure durations.

As for the Chinese learners, in terms of both VOT and closure duration, they confounded Italian voiced stops with the voiceless ones; and produced both categories similarly to Mandarin Chinese unaspirated stops. This confirms our H2 formulated in light of the SLM theory (Flege, 1995, 1996, 2002). It also parallels what was found in our perception experiment, showing that Chinese learners cannot distinguish perceptually between Italian voiced and voiceless stops. Contrary to our H3, the Chinese learners did

not aspirate Italian voiceless [p, t, k]. This suggests that the different orthographic conventions used for stop consonants in Standard Italian and Mandarin Chinese are not a source of confusion for Chinese learners.

In addition, the Chinese learners produced Italian voiceless stops in a native-like fashion. However, consistent with our H4, the Chinese learners' closure durations for Italian voiceless stops were relatively longer than those of the native Italian speakers due to the fact that the learners read at a slower rate than the native speakers, but the differences were not significant. Since in Italian stop closure duration is closely related to stop length (Esposito and Di Benedetto, 1999; Rossetti, 1994), it is likely that the Italian voiceless stops produced by Chinese learners may sound somewhat long to native Italian listeners.

All in all, it can be concluded that Chinese learners approximate the L2 native norms in producing Italian voiceless stops. However, they fail to master Italian voiced stops in production.

V General discussion and conclusions

This study set out to investigate Mandarin Chinese-speaking learners' acquisition of Italian word-initial stop consonants. The results show that Chinese learners have difficulty differentiating perceptually between Italian voiced and voiceless stops; in production, Italian voiced rather than voiceless stops represent a challenge for Chinese learners. The results are in line with the predictions made by the Perceptual Assimilation Model-L2 (PAM-L2) and the Speech Learning Model (SLM), as well as with most other studies focusing on the acquisition of stops of 'true-voice languages' by Chinese learners. In light of this, some considerations are in order.

First, the purpose of the present perception experiment was to compare the category boundaries of the Chinese learners of L2 Italian with those of the native Italian speakers and the native Mandarin speakers. Designing the experiment raises the question of which are the best VOT ranges to use to determine crossover values for speakers differing in their language backgrounds. To eliminate the bias caused by different VOT ranges, we presented all listeners with continua that had identical VOT ranges. However, that being the case, two issues arise. First, since the VOT ranges used did not reflect the actual properties of Italian stops with respect to VOT, it is not easy to ascertain whether the present experimental design has properly activated the 'Italian mode' of the Chinese learners. Thus, further studies employing acoustic stimuli that better reflect the phonetic characteristics of Italian stops may help to consolidate the present conclusions. Second, while the native Mandarin speakers displayed reliable category boundaries, the native Italian participants, being speakers of a true-voice language, diverged from their actual category boundaries due to their sensitivity to 'range effects'. Though this did not have substantial effects on the interpretation of the final results of the present perception experiment, it does imply that for VOT labeling tests that involve speakers with different language backgrounds, participants' different sensitivity to 'range effects' should also be taken into consideration, in addition to the VOT range used.

Second, the category boundaries of the Chinese learners and the native Mandarin speakers for stops of different places of articulation are highly similar to those of native English speakers, which are about 25 ms, 35 ms and 42 ms for bilabial, alveolar and velar

stops respectively (Rojczyk, 2011). According to Keating et al. (1981), these category boundaries are aligned with a natural psycho-acoustic boundary, since both animals (e.g. Dooling, Okanoya & Brown, 1989; Kuhl and Miller, 1975, 1978) and infants (e.g. Eimas et al., 1971; Jusczyk et al., 1989) show similar category boundaries in perception. Therefore, it can be said that these boundaries are inherent in human nature and therefore easily maintained. On the other hand, the perceptual boundary located at 0 ms must be learned in the course of linguistic development (Serniclaes, 2005). For speakers of aspirating languages, generally after six months of age, only the psycho-acoustic boundary remains active (Eilers, Wilson and Moore, 1979), and their increasing language experience tends to enhance this boundary (Werker and Tees, 1984). Thus, it is reasonable to think that for adult Chinese learners it is rather difficult to establish native-like category boundaries for Italian stops. However, as shown by Rochet and Chen (1992), after training, Chinese learners' perceptual identification functions get closer to those of native French speakers. This implies that Chinese learners' ability to differentiate perceptually between Italian voiced and voiceless stops might not easily reach the native level, but could improve via appropriate perceptual training.

Third, the 18 Italian voiced stops realized with prevoicing by the Chinese learners reveal some interesting facts. In the first place, there were fewer [g] realized as voiced than [b] and [d]. This suggests that Chinese learners' acquisition of Italian voiced [g] may occur later than that of [b] and [d]. This parallels the acquisition of Italian voiced stops by L1 Italian infants, who have more difficulty maintaining voicing during velar stops (Bortolini et al., 1995). Thus, in acquiring voiced stops, L1 and L2 learners may follow a similar learning process as the voicing feature is best combined with labiality and worst with velarity (Gamkrelidze, 1975). In the second place, the vast majority of the 18 voiced stops in the data were realized by two Chinese learners. This suggests that these two learners did not produce Italian voiced [b] and [d] by chance. On the contrary, in their productions they systematically differentiated Italian voiced bilabial and alveolar stops from their voiceless counterparts. Moreover, the durations of their prevoicing were rather long. We argue that this is due to hypercorrection. That is, the two Chinese learners were trying to avoid previously recognized errors (i.e. devoicing of Italian voiced stops) through the overproduction of the voicing feature. Since hypercorrection usually occurs at the final stage of acquisition (Eckman, Iverson & Song, 2013), it can be said that these two learners have almost mastered the production of Italian voiced [b] and [d]. Additionally, it should be noted that these two learners were both from the northern dialect area of China, which means they were not familiar with the voicing feature in stops through their dialectal phonology. Though we are unsure how they succeeded in outperforming the other Chinese learners in producing Italian voiced bilabial and alveolar stops, their performance does imply that Chinese learners may acquire Italian voiced stops in production even if they do not perceive them accurately. This adds empirical evidence to what is argued in de Leeuw et al. (2021), showing that accurate L2 production is not necessarily dependent on accurate L2 perception.

Forth, in Italian differences in word meaning are often created using stop voicing distinctions. In addition to the minimal pairs used as stimuli in the production experiment, some examples are *banca* 'bank' vs. *panca* 'bench', *quando* 'when' vs. *quanto* 'how much', *gara* 'race' vs. *cara* 'dear', etc. The inability to perceive and produce the distinction between Italian voiced and voiceless stops will certainly have negative effects on Chinese learners'

language comprehension and intelligibility in real-life communication. As argued before, in production Italian voiced stops are more difficult to acquire than voiceless stops. Thus, Italian voiced stops should be given extra pedagogical attention in teaching activities. Moreover, as voiced stops of different places of articulation present different difficulties for Chinese learners, Italian language instructors should pay attention to the order in which voiced stops are taught, and to the pedagogical efforts dedicated to different voiced stops.

Finally, the present study leaves several crucial issues for future research. For example, this study did not explore the role played by closure duration in the perception of Italian stops by Mandarin Chinese-speaking learners. A future perception experiment with the inclusion of this parameter could help shed light on this issue. Moreover, this study only involved a restricted number of Chinese learners of intermediate level. To better know the development of Chinese learners' acquisition of Italian stop consonants, future research should involve a greater number of Chinese learners varying in their L2 Italian proficiency levels.

Authors' note

Qiang Feng is now affiliated to Dalian University of Foreign Languages, Dalian, China.

Acknowledgements

The authors wish to express sincere gratitude to all the participants involved in this study. Many thanks also to the editors and the three anonymous reviewers for their helpful comments and valuable suggestions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The first author is supported by a grant from the China Scholarship Council (grant number: 201908210329).

ORCID iD

Qiang Feng https://orcid.org/0000-0001-6170-7809

Notes

1. As one of the reviewers pointed out, the actual realization of Standard Italian stop consonants may be subjected to regional diversity. This is because in the different regions of Italy, Standard Italian is spoken with a strong regional inflection. With regard to VOT, it has been shown that in the center and south of Italy, Italian speakers may produce voiceless stops as (partially) voiced (Hualde, Simonet and Nadeu, 2011) or aspirated (Nodari, Celata and Nagy, 2019). Besides, the production of voiceless geminate stops can be accompanied by pre-aspiration (Stevens, 2010). Although beyond the scope of this study, further works addressing the variation in the implementation of Standard Italian stops in different regions are needed.
2. Here the 'orthographic form' refers to the form of 'Pinyin', which is the official Romanization system for Standard Mandarin Chinese in mainland China.

3. The Common European Framework of Reference for Languages (CEFR) is a guideline used to describe foreign language learners' language proficiency. It has six reference levels: from A1 for beginners to C2 for proficient learners. The level of B1 corresponds to an intermediate level.

References

Bates D, Maechler M, Bolker B, et al. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.

Beckman J, Jessen M, and Ringen C (2013) Empirical evidence for laryngeal features: Aspirating vs. true-voice languages. *Journal of Linguistics* 49: 259–84.

Best CT and Tyler M (2007) Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro MJ and Bohn OS (eds) *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins, pp. 13–34.

Boersma P and Weenink D (2020) *Praat: Doing phonetics by computer* [computer program]. Available at: http://www.praat.org (accessed February 2022).

Bortolini U, Zmarich C, Fior R, et al. (1995) Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology* 31: 191–206.

Brady SA and Darwin CJ (1978) Range effect in the perception of voicing. *The Journal of the Acoustical Society of America* 63: 1556–58.

Caramazza A, Yeni-Komshian GH, Zurif EB, et al. (1973) The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *The Journal of the Acoustical Society of America* 54: 421–28.

Chao K and Chen L (2008) A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics and Chinese Language Processing* 2008: 215–32.

Chen Y (2007) *A comparison of Spanish produced by Chinese L2 learners and native speakers: An acoustic phonetics approach*. Urbana, IL: University of Illinois at Urbana-Champaign.

Cho T and Ladefoged P (1999) Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27: 207–29.

Coretta S (2019) An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. *Glossa: A Journal of General Linguistics* 4: 1–25.

de Leeuw E, Stockall L, Lazaridou-Chatzigoga D, et al. (2021) Illusory vowels in Spanish–English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. *Second Language Research* 37: 587–618.

Ding H, Zhan Y, Liao S, et al. (2019) Production of English stops by Mandarin Chinese learners. In: *Proceedings of the 9th International Conference on Speech ProsodyNinth International Conference on Speech Prosody* (pp. 888–892). Poznań, Poland, Shanghai: The International Speech Communication Association (ISCA).

Dooling RJ, Okanoya K, and Brown SD (1989) Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced–voiceless distinction. *Perception and Psychophysics* 46: 65–71.

Duanmu S (2007) *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Eckman FR, Iverson GK, and Song JY (2013) The role of hypercorrection in the acquisition of L2 phonemic contrasts. *Second Language Research* 29: 257–83.

Eilers RE, Wilson WR, and Moore JM (1979) Speech discrimination in the language-innocent and the language-wise: A study in the perception of voice onset time. *Journal of Child Language* 6: 1–18.

Eimas PD, Siqueland ER, Jusczyk P, et al. (1971) Speech perception in infants. *Science* 171: 303–06.

Esposito A (2002) On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59: 197–231.

Esposito A and Di Benedetto MG (1999) Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America* 106: 2051–62.

Flege JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, pp. 233–77.

Flege JE (1996) English vowel productions by Dutch talkers: More evidence for the 'similar' vs. 'new' distinction. In: James A and Leather J (eds) *Second-language speech: Structure and process*. Berlin: Mouton de Gruyter, pp. 11–52.

Flege JE (2002) Interactions between the native and second-language phonetic systems. In: Burmeister P, Piske T, and Rohde A (eds) *An integrated view of language development: Papers in honor of Henning Wode*. Trier: Wissenschaftlicher Verlag, pp. 217–44.

Flege JE and Eefting W (1986) Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica* 43: 155–71.

Flege JE and Eefting W (1987a) Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6: 185–202.

Flege JE and Eefting W (1987b) Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83.

Flege JE and Eefting W (1988) Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America* 83: 729–40.

Fox J and Weisberg S (2019) *An {R} companion to applied regression*. 3rd edition. Thousand Oaks CA: Sage. Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion (accessed February 2022).

Francis AL, Ciocca V, and Ching Yu JM (2003) Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America* 113: 1025–32.

Gabriel C, Kupisch T, and Seoudy J (2016) VOT in French as a foreign language: A production and perception study with mono- and multilingual learners (German/Mandarin–Chinese). In: *Actes du 5e Congrès Mondial de Linguistique Française*. Paris: ILP/EDP Sciences, pp. 1–14.

Gamkrelidze TV (1975) On the correlation of stops and fricatives in a phonological system. *Lingua* 35: 231–61.

Grosjean F (2007) The bilingual's language modes. In: Li W (ed.) *The bilingualism reader*. 2nd edition. Routledge, pp. 428–49.

Hu W (2020) A panel study of Japanese plosive learning by Chinese learners: Taking the perception of *pa* and *ba* as example. *Nihougo no Gakushu to Kenkyu* 1: 61–70.

Hualde JI, Simonet M, and Nadeu M (2011) Consonant lenition and phonological recategorization. *Laboratory Phonology* 2: 301–29.

Jiang X (2020) Zhongguo riyu xuexizhe riyu qingzhuo seyin fayin yanjiu: Yi VOT he GAP wei zhongxin [The production of Japanese voiced and voiceless stops by Chinese learners: Centering on VOT and GAP]. *Comparative Study of Cultural Innovation* 4: 74–76.

Jusczyk PW, Rosner BS, Reed MA, et al. (1989) Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts? *The Journal of the Acoustical Society of America* 85: 1741–49.

Keating PA (1984) Phonetic and phonological representation of stop consonant voicing. *Language* 60: 286–319.

Keating PA, Mikoś MJ, and Ganong III WF (1981) A cross-language study of range of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America* 70: 1261–71.

Kramer M (2009) *The phonology of Italian*. Oxford: Oxford University Press.

Kuhl PK and Miller JD (1975) Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants. *Science* 190: 69–72.

Kuhl PK and Miller JD (1978) Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America* 63: 905–17.

Lennes M (2002) *Calculate_segment_durations*. Praat. Available at: https://github.com/lennes/spect/blob/1.0.0/scripts/calculate_segment_durations.praat (accessed February 2022).

Lenth R (2020) *Emmeans: Estimated marginal means, aka least-squares means*. Available at: https://CRAN.R-project.org/package=emmeans (accessed February 2022).

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

Liu Z and Cebrian J (2016) Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by Mandarin Chinese speakers. Unpublished master's thesis. Universitat Autònoma de Barcelona, Spain.

Liu J, Zeng T, and Lu X (2019) Challenges in multi-language pronunciation teaching: A cross-linguistic study of Chinese students' perception of voiced and voiceless stops. *CÍRCULO de Linguística Aplicada a la Comunicación* 79: 99–118.

Nasukawa K (2005) The representation of laryngeal-source contrasts in Japanese. In: van de Weijer JM, Nanjo K, and Nishihara T (eds) *Voicing in Japanese*. Mouton de Gruyter Berlin, pp. 71–87.

Nearey TM and Rochet BL (1994) Effects of place of articulation and vowel context on VOT production and perception for French and English Stops. *Journal of the International Phonetic Association* 24: 1–18.

Nodari R, Celata C, and Nagy N (2019) Socio-indexical phonetic features in the heritage language context: Voiceless stop aspiration in the Calabrian community in Toronto. *Journal of Phonetics* 73: 91–112.

Peterson RA and Cavanaugh JE (2020) Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics* 47: 2312–27.

Pickett ER, Blumstein SE, and Burton MW (1999) Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica* 56: 135–57.

R Core Team (2020) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, Available at: https://www.R-project.org (accessed February 2022).

Riney TJ, Takagi N, Ota K, et al. (2007) The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics* 35: 439–43.

Ringen C and Kulikov V (2012) Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics* 20: 269–86.

Rochet BL and Chen F (1992) Acquisition of the French VOT contrasts by adult speakers of Mandarin Chinese. In: *Second International Conference on Spoken Language Processing Proceedings*. Edmonton, AB: University of Alberta, pp. 273–76.

Rochet BL and Yanmei F (1991) Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception. *Canadian Acoustics* 19: 105–06.

Rojczyk A (2011) Perception of the English voice onset time continuum by Polish learners. In: Arabski J and Wojtaszek A (eds) *The acquisition of L2 phonology*. Clevedon: Multilingual Matters, pp. 37–58.

Rossetti R (1994) Gemination of Italian stops. *The Journal of the Acoustical Society of America* 95: 2874–74.

Serniclaes W (2005) On the invariance of speech percepts. *ZAS Papers in Linguistics* 40: 177–94.

Shimizu K (1996) *Cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido.

Stevens M (2010) How widespread is preaspiration in Italy? *Working papers*: 97–102.

Sun Y and Profita S (2020) Cross-linguistic study on VOT of Chinese trilingual speakers. *Studies in Literature and Language* 20: 98–102.

Svantesson J-O (1987) Some data on the duration of Chinese stops and affricates. *Working papers/ Lund University, Department of Linguistics and Phonetics* 31: 171–74.

Vagges K, Ferrero FE, Magno-Caldognetto E, et al. (1978) Some acoustic characteristics of Italian consonants. *Journal of Italian Linguistics Amsterdam* 3: 69–84.

Werker JF and Tees RC (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.

Winn MB (2020) Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America* 147: 852–66.

Xu H (2019) L'acquisizione del 'Voice Onset Time' dell'italiano L2 da parte dei sinofoni: Uno studio sperimentale [The acquisition of the 'Voice Onset Time' of Italian L2 by Sinophones: An experimental study]. *H2D − Revista de Humanidades Digitais* 1(1).

Yang Y and Fang Z (1984) Putonghua songqi he busongqi seyin de yinwei jiexian jiqi fanchou zhijue [Phonemic boundary and categorical perception of Mandarin Chinese stops]. In: *Quanguo diwujie xinlixue xueshu huiyi wenzhai xuanji* [Proceedings of the 5th Chinese National Psychology Conference] (pp. 220–29). Beijing: Chinese Psychological Society.

Yang Y, Chen X, and Xiao Q (2022) Cross-linguistic similarity in L2 speech learning: Evidence from the acquisition of Russian stop contrasts by Mandarin speakers. *Second Language Research* 38: 3–29.

Yuan J (2001) *Hanyu Fangyan Gaiyao [Outline of Chinese dialects]*. Beijing: Yuwen chubanshe.

Zhang J (2012) Zhongguo xuesheng xide fayu seyin de fenxi [Analysis of Chinese students' production of French Stops]. *International Journal of Chinese Studies* 2: 174–81.

Zhang J (2014) Stops consonants in Mandarin in the perspective of perception. *Journal of Yunnan Normal University (Teaching and research on Chinese as a foreign language edition)* 12: 58–64.

Zhang S (2013) Zhongguo xuesheng fayu xuexi zhong busongqi qingseyin yu zhuoseyin de wenti [Chinese students' problems in the acquisition of French unaspirated and voiced stops]. *Theory and Practice of Contemporary Education* 5: 100–02.

Appendix 1. Word stimuli for the production experiment.

Language	stop	Target words		Distractors	
Standard Italian	[b]	*batto* 'I hit'	*basso* 'short'	*rana*	*lana*
	[p]	*patto* 'pact'	*passo* 'step'	*rotto*	*lotto*
	[d]	*dare* 'give'	*data* 'date'	*russo*	*lusso*
	[t]	*tale* 'this'	*tata* 'nanny'	*regno*	*legno*
	[g]	*gatto* 'cat'	*gallo* 'rooster'	*sette*	*sete*
	[k]	*cassa* 'cash desk'	*callo* 'callus'	*Lucca*	*Luca*
				pappa	*papa*
				pollo	*polo*
Mandarin Chinese	[p]	<巴士> 'bus'	<掰开> 'break apart'	<来了>	<奶奶>
	[p^h]	<趴着> 'lie down'	<拍照> 'take photo'	<扎实>	<摘抄>
	[t]	<搭车> 'take car'	<呆板> 'stiff'	<拉手>	<哪里>
	[t^h]	<他的> 'his'	<胎儿> 'fetus'	<歪了>	<挖掘>
	[k]	<嘎嘎> 'quack'	<该国> 'this country'	<牛奶>	<流浪>
	[k^h]	<咖啡> 'coffee'	<开门> 'open door'	<森林>	<申辩>
				<四十>	<十四>
				<手机>	<搜罗>