UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Linguistic and Literary Studies

Ph.D. COURSE IN: Linguistic, Philological and Literary Sciences
SERIES XXXV

The Acquisition of Italian Consonant Sounds by Mandarin Chinese-Speaking Learners

Thesis written with the financial contribution of China Scholarship Council

Coordinator: Prof. Rocco Coronato
Supervisor: Prof.ssa Maria Grazia Busà

Ph.D. student: Qiang Feng

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor Prof. Dr. Maria Grazia Busà. We are so compatible not only academically, but also in character. The past three years we spent together are marked by countless emails, meetings, discussions, and also laughter between us. It is her expertise, patience, and motivation that help me complete this dissertation.

Also, I'm extremely grateful to all my participants (most of them are my dear students from Dalian University of Foreign Languages). Without their participation in my experiments, this dissertation would not have been possible.

Special thanks to my ten-year friend Dr. Peng Li, not only for his help with statistics during the writing of this dissertation, but also for the invaluable and wise advice he gave me at the important moments of my life.

I am very thankful to my friends in Padova: Yi Yang, Shaomin Xie, and Luca Zhou. I will always remember the time we spent together studying, cooking, eating, drinking, talking, shopping, and traveling. It is their company that has made me feel less alone in the past three years.

Thanks should also go to my friends in China: Xudong Huo, Jia Wang, Bin Zhang, Zhexun Fu, Jinfei Li, Shu Jiang, Yifan Qian, Xulong Wu, and Yusheng Yang. It is our lighthearted conversations on WeChat that give me a break from my research work every now and then.

Lastly, I'd like to mention my dearest parents, sister, and little niece. They are always my spiritual support and the motivation for my hard work.

ABSTRACT

The acquisition of Italian consonant sounds by Mandarin Chinese-speaking learners has hardly been empirically investigated. This doctoral dissertation aims to fill this gap by investigating three areas which have been documented to be problematic for Chinese learners, that is, how Chinese learners acquire Italian voiced vs. voiceless stop consonant contrast, singleton vs. geminate consonant contrast, and lateral vs. rhotic consonant contrast.

In the first study, twenty Mandarin Chinese-speaking undergraduate students majoring in Italian, five native Italian and five native Mandarin speakers served as participants in a perception experiment; and an equal number of participants with the same language backgrounds served as participants in a 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 −50ms to 90ms in 10ms 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 second study had ten first-year, ten second-year and ten third-year Chinese undergraduate students majoring in Italian and ten native Italian-speaking controls as participants. In the perception experiment, the participants had to identify ten Italian disyllabic minimal pairs contrasting in consonant length. In the production experiment, the participants were asked to read five out of the ten minimal pairs mentioned above; the duration values of the intervocalic consonants and the preconsonantal vowels were measured and converted into duration ratios for statistical analyses. The results show that, in both perception and production, Chinese learners can distinguish between the two consonant length categories in Italian to a certain extent, but not in a native-like manner. Also, the duration interplay between Italian consonants and preconsonantal vowels (i.e., longer vowels before singleton consonants and shorter vowels before geminate consonants) is entirely ignored by Chinese learners. Moreover, Chinese learners' increased learning experience does not appear to enhance their acquisition of Italian consonant length contrast.

The participants of the third study were thirty Chinese learners of Italian with different learning experiences and ten native Italian speakers. In the perception experiment, the participants identified six Italian minimal pairs contrasting in /r-l/. In the production experiment, the participants read the six minimal pairs, and their productions were assessed by three native Italian-speaking raters. The results show that, in perception, Chinese learners have some difficulty differentiating between Italian /r-l/ contrast. In production, Chinese learners have more difficulty properly realizing Italian /r/ than /l/, and show the tendency to replace /r/ with /l/. Also, while Chinese learners' production of Italian /r-l/ contrast varies with their increased learning experience, their perceptual accuracy remains unchanged.

All in all, this doctoral dissertation provides a detailed picture of how Chinese learners acquire the stop contrast, consonant length contrast, and /r-l/ contrast in Italian, showing that the three consonant contrasts pose acquisition difficulties for Mandarin Chinese-speaking learners. Moreover, the plausible reasons for these acquisition difficulties are discussed.

Keywords: L2 speech acquisition, consonant sounds, Italian, Mandarin Chinese

CONTENTS

Chapter 1

General Introduction

1.1 Mandarin Chinese-speaking learners of Italian

Italian is one of the most learned languages in the world. According to the latest report on the general status of the Italian language in the world (i.e., *L'italiano nel Mondo che Cambia - 2019* 'Italian in the Changing World - 2019') released by Italian MAECI (*Ministero degli Affari Esteri e della Cooperazione Internazionale* 'Ministry of Foreign Affairs and International Cooperation'), in the 2017/2018 academic year, outside Italy, there were more than two million learners of Italian in 119 countries. Out of these countries, China ranked 29th with 10,520 learners. As far as Asia is concerned, China ranked second only after Japan. Most of the 10,520 Mandarin Chinese-speaking learners of Italian were university students (about 38.2% of the total, equal to 4,018 learners). According to the data provided by *Istituto Italiano di Cultura - Pechino* 'Italian Cultural Institute - Beijing', the first time that the Italian language became an independent major in the Chinese university system was 1954 at the University of International Business and Economics in Beijing; and as of 2021, there are 24 Chinese universities (in 16 cities) that have the Italian language as an independent major.

Inside Italy, the position occupied by Mandarin Chinese-speaking learners is even more prominent as compared to that outside Italy. According to the statistical data of the 'Global Flow of Tertiary-Level Students' of UNESCO (http://uis.unesco.org/en/uis-student-flow), for Chinese students studying abroad, Italy is the 11th most popular destination. Moreover, among the international students coming to study in Italy, Chinese students (11,965) rank first, almost four times the number of Indian students (3,021) that rank second.

According to the data of Uni-Italia (*Centro di Promozione Accademica per l'Orientamento allo Studio in Italia* 'Centre for the Academic Promotion and Orientation of Study in Italy'), in the 2018/2019 academic year, among the foreign students enrolled in the public and private universities of Italy, Chinese students

(7,376) ranked second only after Romanian ones (9,630). However, as stated by Uni-Italia, since a large number of Romanian students were already Italian residents, this rank could be very different if they were excluded from the database. As for the AFAM (*Alta Formazione Artistica e Musicale* 'Higher Education in Art and Music') institutions of Italy (e.g., *Accademie di belle arti* 'Art schools', *Conservatori di musica* 'Music schools', etc.), the predominance of Chinese students is even more striking. Specifically, in the 2018/2019 academic year, more than half of the international students studying in Italian AFAM institutions were Chinese students (7,485 Chinese students/12,103 international students = 61.8%).

The predominance of Chinese students among international students studying in Italy can be attributed to two programs. That is, the Marco Polo program that started in 2005, and the Turandot program that started in 2009. These two programs are government agreements signed between Italy and China, aiming at facilitating the presence of Chinese students in Italian universities (Marco Polo) and AFAM institutions (Turandot). Through these two programs, even the Chinese students without any previous knowledge of the Italian language can be pre-enrolled in a university or AFAM course. However, before the official enrolment, it is compulsory for the Chinese students to take an Italian language course in Italy that lasts 10-11 months. At the end of the course, the students are expected to reach at least the B1 level as defined by the Common European Framework of Reference for Languages (CEFR), and to increase their language skills in the following years. According to the data provided by Uni-Italia, in the 2020/2021 academic year, there were almost 3,000 Chinese students (880 via Marco Polo and 2,090 via Turandot) coming to study in Italy despite the Covid-19 pandemic.

Considering such a large number of Mandarin Chinese-speaking learners of Italian both outside and inside Italy, it is rather surprising to find that very few empirical studies have been conducted to examine how Mandarin Chinese-speaking learners acquire L2 Italian speech. Among the very limited number of studies focusing on this acquisition process, the major ones are reported in the following section.

1.2 Major studies on the acquisition of L2 Italian speech by Mandarin Chinese-speaking learners

De Meo et al. (2015) investigated the developmental acquisition of Italian vocalic system by Mandarin Chinese-speaking learners of different language proficiencies. In their study, the Chinese learners were ask two read two texts in Italian and two in Mandarin Chinese, which contained the entire phonetic-phonological inventory of Italian and the entire vocalic repertory of Mandarin Chinese, respectively. The $F1$, $F2$ and duration values of the learners' vowel productions were analyzed. The results show that, first, Chinese students fail to differentiate Italian stressed and unstressed vowels in production. This is because they produce the two types of vowels with similar duration, rather than producing stressed vowels that are longer than the corresponding unstressed ones, as native Italian speakers do. Second, Chinese learners produce Italian /a, i, u/ well, but have difficulty producing Italian mid-open (i.e., /ɛ/ and /ɔ/) and mid-closed vowels (i.e., /e/ and /o/) distinctively. Moreover, this difficulty does not seem to decrease with Chinese learners' increased proficiency.

As for prosody acquisition, De Meo and Pettorino (2010) examined how Mandarin Chinese-speaking learners perceive and produce three speech acts in Italian, namely command, request and concession. In their study, the Chinese participants first identified perceptually the three Italian speech acts produced by L1-Italian speakers. Their overall perceptual identification accuracy reached 59%, which was lower than the accuracy of L1-Slavic learners (77%), but higher than that of L1-Vietnamese learners (48%). Afterwards, the Chinese participants produced the three Italian speech acts, and their productions were assessed perceptually by native Italian listeners. The results show that, overall, the Chinese learners' prosodic-pragmatic realization was recognized correctly only 33% of the times, with command and concession having the highest (38%-49%) and lowest (23%-25%) recognition accuracy, respectively. Moreover, the increased language competence played only a marginally positive role in improving the Mandarin Chinese-speaking learners' acquisition of Italian prosody, indicating a possible lack of pedagogical attention paid to the suprasegmental aspects of language communication.

Regarding Mandarin Chinese-speaking learners' acquisition of Italian consonant sounds, however, most investigations are descriptive studies based on the authors' impressionistic observations from either teaching experiences (e.g., Calabrò & Mairano, 2017; Costamagna, 2010; Trifone, 2014; Zhou, 2012) or conversational interactions (e.g., Dalla Palma, 2019). The remaining handful of empirical studies (Costamagna et al., 2014; Sun & Profita, 2020; Xu, 2019) have some common limitations, such as a lack of analyses of perceptual data, no support of rigorous statistical analyses, and small sample size (for detailed reviews of these studies, see the following relevant chapters). Thus, it can be said that the acquisition of L2 Italian consonants by L1-Mandarin Chinese learners remains a much understudied field.

To fill this gap, the aim of this dissertation is to shed more light on Mandarin Chinese-speaking learners' acquisition mechanisms of Italian consonant sounds by providing a more rigorous scientific investigation of this issue. To this end, a comparison of Standard Italian and Mandarin Chinese consonant systems will first be made.

1.3 A comparison of Standard Italian and Mandarin Chinese consonant inventories

Table 1 and Table 2 show Italian consonant sounds alongside Mandarin Chinese ones for comparison. Some differences (and similarities) that can be noticed directly from this comparison are as follows.

Table 1. Standard Italian consonant phonemes (adapted from Rogers & d'Arcangeli, 2004, p. 117)

	Bilabial	Labiodental	Alveolar	Post-alveolar	Palatal	Velar	Labial Velar
Stop	p b		t d			k g	
Affricate			t͡s d͡z	t͡ʃ d͡ʒ			
Fricative		f v	s z	ʃ			
Nasal	m		n		ɲ		
Lateral			l		ʎ		
Approximant					j		w
Trill			r				

Table 2. Mandarin Chinese consonant phonemes (adapted from Lee & Zee, 2003, p. 109)

	Bilabial	Labiodental	Dental/Alveolar	Post-alveolar	Palatal	Velar
Stop	p pʰ		t tʰ			k kʰ
Affricate			t͡s t͡sʰ	t͡ʃ t͡ʃʰ	tɕ tɕʰ	
Fricative		f	s	ʃ	ɕ	x
Nasal	m		n			ŋ
Lateral			l			
Approximant	w			ɹ	j	

First, as for stop consonants, both Standard Italian and Mandarin Chinese have three pairs of stop contrasts, and their places of articulation, namely bilabial, alveolar and velar, are identical in these two languages. However, the stop contrasts in Standard Italian and Mandarin Chinese are distinguished by different features: Standard Italian stops are distinguished by voicing, and can be subdivided into voiced /b, d, g/ and voiceless /p, t, k/. On the other hand, all stops in Mandarin Chinese are voiceless, and are distinguished by aspiration. That is, Mandarin Chinese stops can be subdivided into aspirated /p^h, t^h, k^h/ and unaspirated /p, t, k/. So, Standard Italian and Mandarin Chinese share the voiceless unaspirated /p, t, k/, but their voiced counterparts, namely /b, d, g/, are present in the Italian consonant inventory but are absent from the Mandarin Chinese phonology; and in the Mandarin Chinese consonant system, in the place of voiced /b, d, g/, there are voiceless aspirated /p^h, t^h, k^h/.

Second, as for affricates, Mandarin Chinese has one more contrast than Standard Italian, namely the palatal /tɕ/ vs. /tɕ^h/. The other two pairs of affricate contrasts in these two languages have the same places of articulation, namely alveolar (Italian /t͡s/ vs. /d͡z/ and Mandarin /t͡s/ vs. /t͡s^h/) and post-alveolar (Italian /t͡ʃ/ vs. /d͡ʒ/ and Mandarin /t͡ʃ/ vs. /t͡ʃ^h/). Similar to stop consonants, the affricate contrasts in Standard Italian are distinguished by voicing, whereas in Mandarin Chinese they are distinguished by aspiration. Thus, Mandarin Chinese shares /t͡s/ and /t͡ʃ/ with Standard Italian, yet the voiced counterparts of these two phonemes, namley /d͡z/ and /d͡ʒ/, exist in Italian but not in Mandarin Chinese; and in the Mandarin Chinese phonological system, in the place of voiced /d͡z/ and /d͡ʒ/ are voiceless aspirated /t͡s^h/ and /t͡ʃ^h/.

Third, concerning fricative consonants, the labiodental /f/, alveolar /s/ and post-alveolar /ʃ/ exist in both Standard Italian and Mandarin Chinese. Again, the voiced counterparts of the voiceless /f/ and /s/ that appear in Standard Italian, namely /v/ and /z/, are absent from the Mandarin Chinese consonant system. Besides, Mandarin

Chinese has two fricatives that do not appear in Italian phonology: the palatal /ɕ/ and the velar /x/.

Fourth, both Standard Italian and Mandarin Chinese have three nasal consonants, of which the bilabial /m/ and the alveolar /n/ are common to these two languages. The nasal that is present in Standard Italian but not in Mandarin Chinese is the palatal /ɲ/, and the nasal that is present in Mandarin Chinese but not in Standard Italian is the velar /ŋ/.

Fifth, Mandarin Chinese phonology has only one lateral consonant, namely the dental/alveolar /l/, which exists also in Standard Italian. The palatal lateral /ʎ/ appears only in Standard Italian but not in Mandarin Chinese.

Sixth, as for approximant consonants, the only one that is common to both the Standard Italian and Mandarin Chinese phonology is the palatal /j/. As for /w/, its place of articulation differs to some extent in Standard Italian (labial velar) and Mandarin Chinese (bilabial). Besides, Mandarin Chinese has the post-alveolar /ɹ/, which is absent in Standard Italian.

Seventh, the Mandarin Chinese phonological system lacks trill consonants. By contrast, Standard Italian has the trill /r/, which can be realized with different allophonic variants across different phonetic contexts. Specifically, Italian word-initial /r/ is usually realized as a trill [r] (Bertinetto & Loporcaro, 2005; Celata et al., 2016), while the (singleton) intervocalic /r/ is often reduced to a tap [ɾ] (Bertinetto & Loporcaro, 2005; Kramer, 2009; Rogers & d'Arcangeli, 2004). In other positions, Italian /r/ is generally pronounced as a trill [r] in stressed syllables and as a tap [ɾ] in unstressed ones (Canepari, 1999). Besides, other individual variants (e.g. uvular trill, alveolar approximant and fricative) may also be encountered (Celata et al., 2016; Romano, 2013).

In addition to the above differences, there is another one that does not appear from the comparison of the two tables. That is, the stop (/p b t d k g/), affricate (/t͡ʃ d͡ʒ/), fricative (/f v s/), nasal (/m n/), lateral (/l/) and trill (/r/) consonants in Standard Italian can be geminated (for stops, see Esposito & Di Benedetto, 1999; for affricates and fricatives, see Di Benedetto & De Nardis, 2021a; for nasals, lateral and trill, see Di Benedetto & De Nardis, 2021b). In terms of duration, Italian geminate consonants are longer than their singleton counterparts and they contrast in intervocalic position (e.g., *nono* 'ninth' vs. *nonno* 'grandfather', *caro* 'dear' vs. *carro* 'wagon'). Moreover, in Italian minimal pairs contrasting in singleton/geminate consonants, the duration of pre-geminate vowel is shorter than that of the corresponding pre-singleton vowel. On the other hand, in Mandarin Chinese there is no consonant (and vowel) length distinction at the phonemic level.

Regarding consonant phonotactics, in Mandarin Chinese, consonants cannot occur in syllable coda position, except for the dental/alveolar nasal /n/ and the velar nasal /ŋ/. Besides, Mandarin Chinese has no consonant clusters within a single syllable (Duanmu, 2007). By contrast, almost all Standard Italian consonants are allowed in syllable coda position; and Standard Italian allows consonant clusters consisting of up to three consonants in syllable-initial position (Kramer, 2009; Marotta & Vanelli, 2021).

To sum up, Standard Italian and Mandarin Chinese share a certain amount of consonant sounds (i.e. the stops /p, t, k/, the affricates /t͡s, t͡ʃ/, the fricatives /f, s, ʃ/, the nasals /m, n/, the lateral /l/ and the approximant /j/). Also, the approximant /w/ in these two languages is somewhat similar. Nevertheless, there are clear differences between the consonant systems of the two languages. First, many consonants are not shared by Standard Italian and Mandarin Chinese. Specifically, the consonants that exist in Standard Italian but not in Mandarin Chinese are the stops /b, d, g/, the affricates /d͡z, d͡ʒ/, the fricatives /v z/, the nasal /ɲ/, the lateral /ʎ/, and the trill /r/. The consonants that are present in Mandarin Chinese but absent in Standard Italian include the stops /pʰ, tʰ, kʰ/, the affricates /tɕ, tɕʰ, t͡sʰ, t͡ʃʰ /, the fricatives /ɕ, x/, the nasal /ŋ/ and the approximant /ɹ/. Second, the consonant length contrast that is present in Standard Italian is absent in Mandarin Chinese. Third, compared with Mandarin Chinese, Standard Italian has a much larger number of consonants occurring in syllable coda position; and the only syllable coda consonant that is allowed by both Standard Italian and Mandarin Chinese is the nasal /n/. Forth, consonant clusters exist only in Standard Italian but not in Mandarin Chinese.

1.4 Consonant problems for Mandarin Chinese-speaking learners of Italian

It is generally believed that for L2 learners, the L2 phonemes that are available also in their L1 phonology usually represent no acquisition challenge; by contrast, the L2 phonemes that are absent from their L1 phonology can be hard to master. Therefore, in our case, Mandarin Chinese-speaking learners should have no difficulty acquiring Italian syllable-initial /p, t, k, t͡s, t͡ʃ, f, s, ʃ, m, n, l, j/, which exist also in Mandarin Chinese. As for Italian /w/, Mandarin Chinese-speaking learners should have some difficulty with it for its being similar but not identical to Mandarin Chinese /w/. The main potential consonant problems encountered by Mandarin Chinese-speaking learners of Italian should be how to accurately perceive and produce Italian /b, d, g, d͡z, d͡ʒ, v, z, ɲ, ʎ, r/ and consonant length contrast, which are completely absent from the Mandarin Chinese phonological system. Also, the syllable-final consonants and consonant clusters that are present in the Standard Italian but not Mandarin Chinese phonology should be challenging for L1-Chinese learners.

However, what are the *actual* rather than *potential* problems that Mandarin Chinese-speaking learners have in their Italian consonant learning process? To answer this question, we referred to the descriptive studies mentioned in § 1.2, which, to our knowledge, are the only ones that are available on this issue. In these studies, the authors (who are usually native Italian-speaking language instructors or scholars) reported their personal observations of the recurring consonant errors that Mandarin Chinese-speaking learners make in their L2 Italian classrooms or real-life communications. Therefore, though these studies are descriptive rather than scientific, and so their conclusions are based on intuitions rather than rigorous empirical procedures, they faithfully document the actual problems that Mandarin Chinese-speaking learners have in their Italian consonant learning process.

In reviewing these studies, we found three common consonant problems that have been documented in nearly all studies. These regard the acquisition of Italian voiced/voiceless stop contrast, singleton/geminate consonant contrast, and /r-l/ contrast. The detailed descriptions of these consonant problems are as follows.

First, regarding stop consonants, it has been noted that Mandarin Chinese-speaking learners have difficulty differentiating between Italian voiced and voiceless stops in both perception (Costamagna, 2010; Zhou, 2012) and production (Costamagna, 2010; Dalla Palma, 2019; Trifone, 2014; Valentini, 1990). For example, Mandarin Chinese-speaking learners may produce Italian *denti* 'teeth' as *tenti**, and *campo* 'field' as *cambo**. Second, as for consonant length contrast, Mandarin Chinese-speaking learners' main problem is that they tend to degeminate Italian geminate consonants in production (Costamagna, 2010; Dalla Palma, 2019; Valentini, 1990; Zhou, 2012). For example, they may produce *sorella* 'sister' as *sorela**. Third, the most notable feature in the Italian speech by Mandarin Chinese-speaking learners concerns the /r-l/ contrast, and specifically the tendency to confound these two phonemes in both perception (Costamagna, 2010) and production (Costamagna, 2010; Dalla Palma, 2019; D'Annunzio, 2009; Trifone, 2014; Valentini, 1990; Zhou, 2012). For instance, *mare* 'sea' may be produced as *male* 'bad', and *loro* 'them' as *rolo** by Mandarin Chinese speaking-learners.

It is interesting to note that each of the above three types of contrast has a high functional load (Martinet, 1952): that is to say, each of them differentiates a great number of minimal pairs in Italian (Mairano & Calabrò, 2017). Therefore, the problematic perception and production of these contrasts may significantly affect the speech comprehension and intelligibility of Mandarin Chinese-speaking learners of Italian. Thus, these problems appear to be well worth investigating in an in-depth way. This is because these investigations may not only give insights into Mandarin Chinese-speaking learners' L2 consonant acquisition mechanisms, but also provide some pedagogical implications on how to aid Mandarin Chinese-speaking learners to overcome these acquisition challenges.

Given that these consonant problems have hardly been empirically investigated, many detailed issues remain open. For example: What is the exact way in which Mandarin Chinese-speaking learners' perceive and produce these contrasts? How do their perception and production relate to each other? Do Mandarin Chinese-speaking learners with longer learning experience outperform the less experienced ones? Most importantly, what are the plausible reasons for these acquisition difficulties. Are there any possible pedagogical solutions to these difficulties? To shed some light on these questions, the present dissertation includes three empirical studies that address respectively how Mandarin Chinese-speaking learners acquire Italian voiced/voiceless stop contrast, singleton/geminate consonant contrast, and /r-l/ contrast.

1.5 Structure of the dissertation

This dissertation consists of five chapters, which are structured as follows.

In Chapter 2 we will look at how Mandarin Chinese-speaking learners perceive and produce Italian word-initial stops in light of the Perceptual Assimilation Model-L2 (PAM-L2; Best & Tyler, 2007) and the Speech Learning Model (SLM; Flege, 1995). Both models are widely tested and assume that it is the different degrees of (dis)similarity between L1 and L2 sounds that determine the different degrees of success in L2 speech learning. In short, according to PAM-L2, if two L2 sounds are perceptually assimilated to two different L1 sound categories (Two Category assimilation), L2 learners will have high discrimination accuracy. By contrast, in the case of Single Category assimilation (i.e., two L2 sounds are assimilated to the same L1 sound category), L2 learners will discriminate them poorly. In other cases (i.e., Category Goodness assimilation, Uncategorized-Categorized assimilation, Uncategorized-Uncategorized assimilation, and Non-Assimilable case), L2 learners' perceptual discrimination performance may range from poor to good. As for L2 production, according to SLM, it is closely related to L2 perception. Specifically, only when an L2 sound is accurately perceived, can it be (possibly) accurately realized in production. On the contrary, if inappropriate operation of equivalence classification (i.e., inappropriate assimilation of L2 sounds to the close L1 counterparts) takes place, L2 learners will have problems with the production of target L2 sounds. Under these theoretical premises, we will look at the (dis)similarity between Standard Italian and Mandarin Chinese stops and investigate how this (dis)similarity affects Mandarin Chinese-speaking learners' acquisition of Italian stop consonants.

Chapter 3 will examine how Mandarin Chinese-speaking learners acquire Italian consonant length contrast following the Second Language Linguistic Perception Model (L2LP; Escudero, 2005) and the Feature Model (Brown, 1998), as these two models (and also the revised version of L2LP model [Van Leussen & Escudero, 2015]) take into account duration properties in predicting L2 speech acquisition. According to the L2LP model, at the initial stage of L2 perceptual learning, L2 learners tend to duplicate their L1 grammar in their L2 perception. However, through increased exposure to the L2 perceptual input, L2 learners will modify their L2 perception grammar and ultimately become native-like. As for production, according to the Feature Model, in the case of L2 consonant length contrast acquisition, if the feature [± long] is exploited in L2 learners' native language, it will be transferred and reused in L2 learning and will thereby favor L2 learners' production accuracy. On the contrary, L2 learners that are not familiar with the feature [± long] through their L1 phonology will not have this inherent advantage. In our case, consonant length contrast is absent from the Mandarin Chinese phonology; thus, we aim to verify whether Chinese learners follow the predictions of the two models above in the acquisition of Italian singleton/geminate consonant contrast.

Chapter 4 will investigate the perception and production of Italian /r-l/ contrast by Mandarin Chinese-speaking learners with different learning experiences in light of a recent language acquisition theory, namely the Perceptual Assimilation Model combined with the Articulatory Organ Hypothesis (PAM-AOH; Best et al., 2016). PAM-AOH divides consonant contrasts into three types, namely *between-organ* distinctions (i.e., two consonants are realized with different articulatory organs), *within-organ* distinctions (i.e., two consonants are realized with the same articulatory organ), and *privative organ* distinctions (i.e., a specified gesture of a certain articulatory organ is present in one consonant but absent in the other). The between-organ distinctions are easily discriminated. By contrast, the perceptual discrimination of within-organ and privative organ distinctions can be difficult if they are nonnative. As for production, PAM-AOH posits that it closely correlates with perception. Specifically, it is the perceived articulatory information that guides speech production. Following these premises, we will take a close look at the articulatory properties of Italian /r-l/ contrast, and examine the extent to which PAM-AOH can account for the acquisition of this consonant contrast by Mandarin Chinese-speaking learners.

In Chapter 5, we will summarize the main findings and implications obtained from the above three empirical studies, point out their limitations, and indicate some future research directions.

Chapter 2

Acquiring Italian Stop Consonants: A Challenge for Mandarin Chinese-Speaking Learners

[Feng, Q., & Busà, M. G. (2022). Acquiring Italian stop consonants: A challenge for Mandarin Chinese-speaking learners. *Second Language Research*. https://doi.org/10.1177/02676583221079147]

2.1 Introduction

The acquisition of stop consonants is a crucial issue in foreign language learning for Mandarin Chinese speakers. This has been investigated in various L2 or 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 et al. (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 et al., 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 et al., 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 et al., 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.

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.

2.2 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,

[1] 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 et al., 2011) or aspirated (Nodari et al., 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.

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 3. 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.

Table 3. 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 and Chen (2008)	14	16	27	82	81	92

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.

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 (see Table 4 for a comparison). 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

[2] Here the “orthographic form” refers to the form of “Pinyin”, which is the official romanization system for Standard Mandarin Chinese in mainland China.

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].

Table 4. 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>[3]			
	Chinese				<b>	<d>	<g>	<p>	<t>	<k>

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.

[3] Note that Italian <ch>, <q> and <k> are also pronounced [k].

Thirdly, the closure durations of Italian stop consonants have been found to be inversely related to speaking rates (Pickett et al., 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.

2.3 Perception experiment

2.3.1 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. At the time of the experiment, none of them had had a study abroad experience in Italy. 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)[4] according to the assessment of their native Italian-speaking teachers.

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 et al. (1981) examined a series of alveolar continua with different VOT ranges, and found that the perceptual boundaries of the same listeners

[4] 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.

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 created following the tutorial in Winn (2020), which generates VOT continua through the manipulation of aspirated and unaspirated stops. Since both types of stops are present in the Mandarin Chinese (but not Italian) phonology, a native Mandarin Chinese speaker from the northern dialect area of China was recruited for the original sound recording. He was instructed to produce three pairs of monosyllables, namely [pa] vs [p^ha], [ta] vs [t^ha], and [ka] vs [k^ha] 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 ExperimentMFC 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; no test stimuli were included) 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 4050 responses (135 stimuli × [20+5+5] participants = 4050). This low percentage of misperceptions (238/4050 = 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 × 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 4]. 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).

2.3.2 Results

In total, we had 3812 valid responses (4050 responses – 238 misperceptions = 3812). 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.

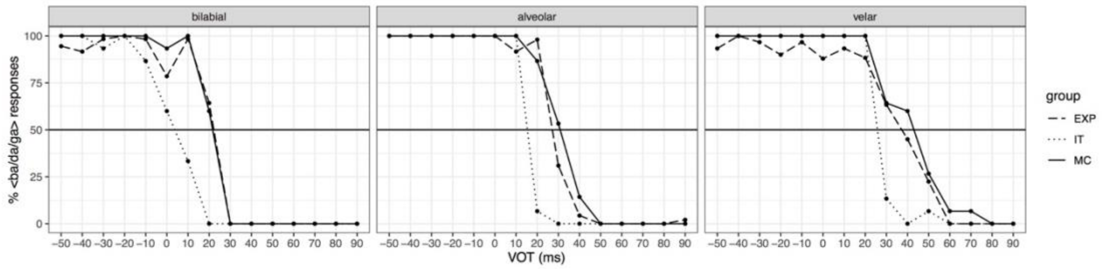


Figure 1. The EXP, IT and MC groups' average labeling functions for the bilabial, alveolar and velar VOT continua.

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 5. 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.

Table 5. Summary of the results of the comparisons of contrasts at the EXP group's near-crossover values for the bilabial, alveolar and velar continua (p values < 0.05 are in bold).

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**

2.3.3 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' 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 (cf. Table 3), 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.

2.4 Production experiment

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

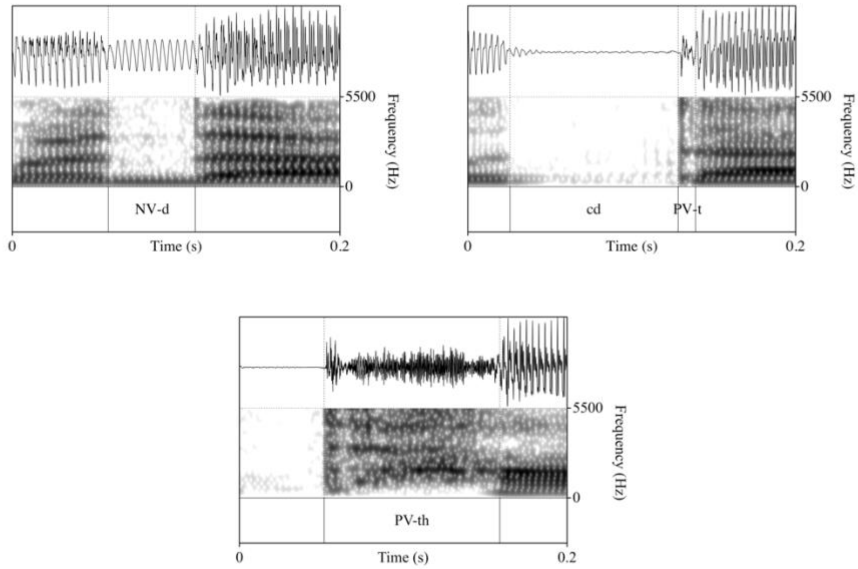


Figure 2. Acoustic waveforms and spectrograms at a 200 ms time scale of (i) the VOT of an Italian voiced [d] produced by the participant IT-2 in the IT group (upper left panel); (ii) the closure duration and VOT of an Italian voiceless unaspirated [t] produced by the participant EXP-20 in the EXP group (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. NV = negative VOT. PV = positive VOT. cd = closure duration.

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 4). 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).

2.4.2 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 6, 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 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: −147.5ms).

Table 6. 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%

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.

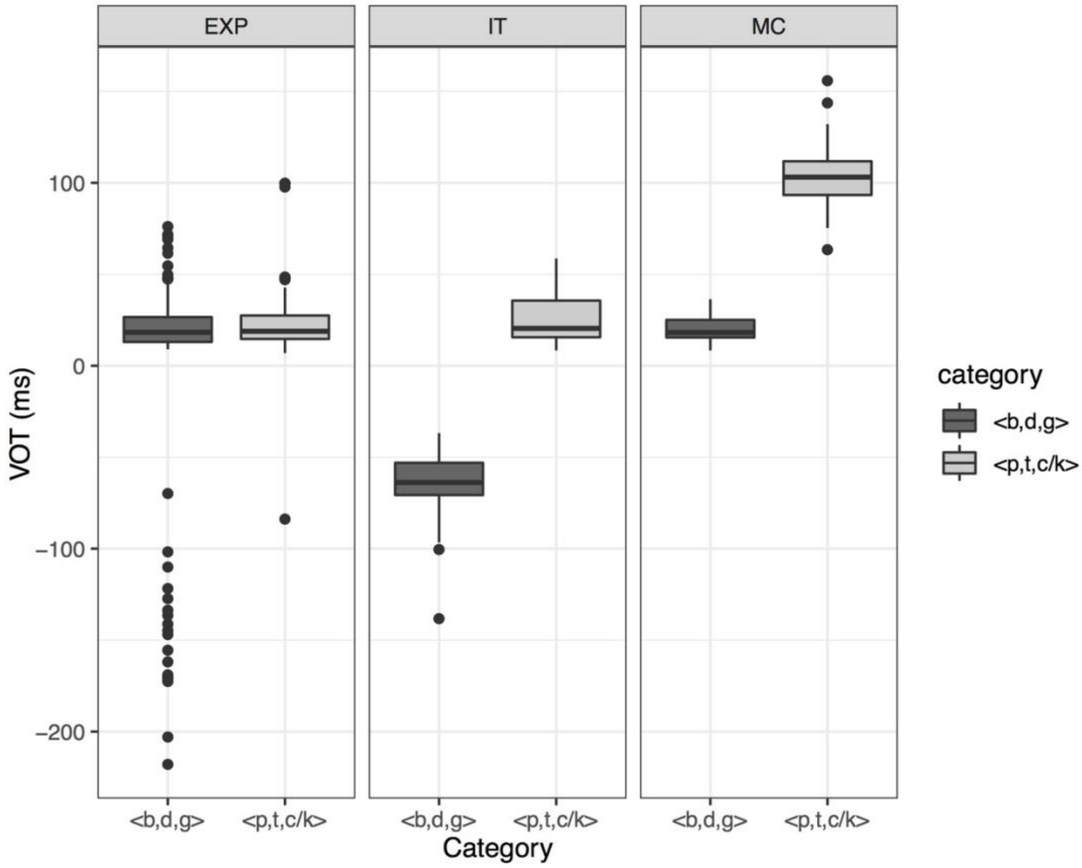


Figure 3. VOT distributions of the stops produced by the EXP, IT and MC groups.

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 showed 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 7, 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 7. 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)

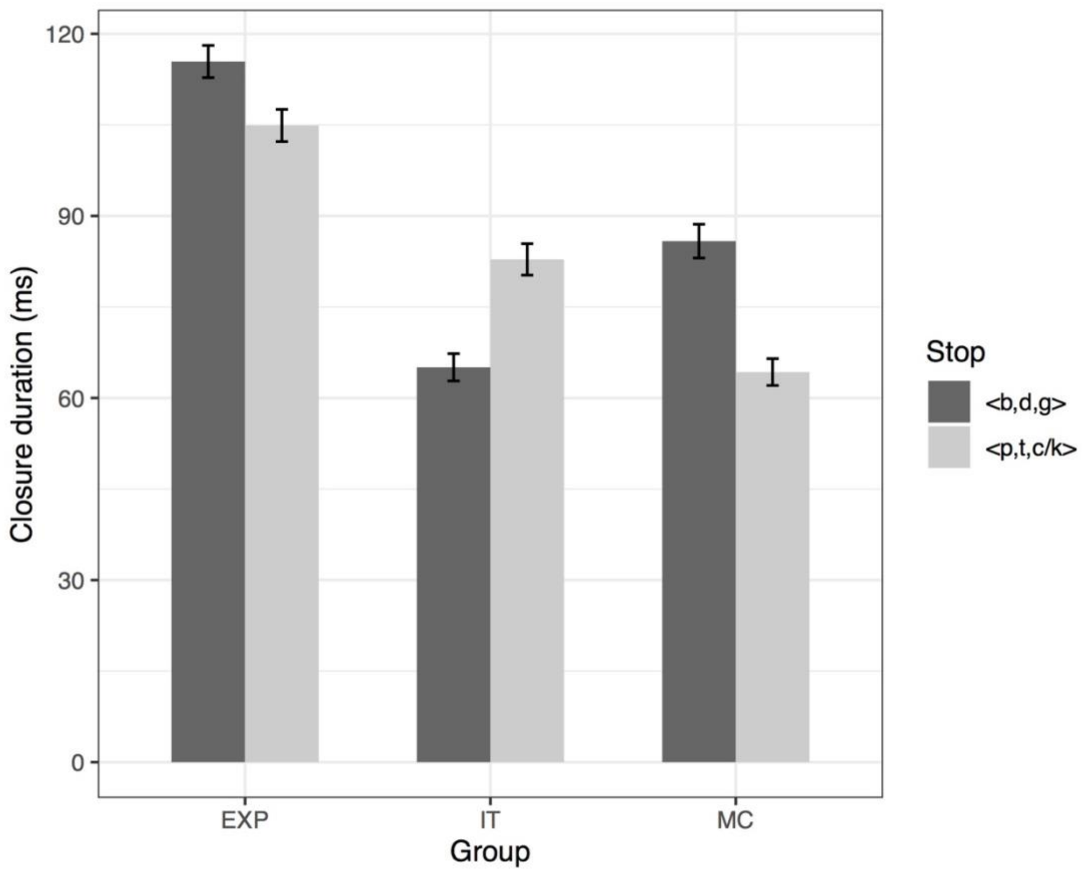


Figure 4. Average closure durations of the stops produced by the EXP, IT and MC groups.

For the statistical analyses, the closure duration values were first normalized using the bestNormalize 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 (*β* = –0.68 ± 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 (*β* = 0.82 ± 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> (*β* = 0.28 ± 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 (*β* = 1.48 ± 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 (*β* = 0.52 ± 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> (*β* = 0.69 ± 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> (*β* = –0.41 ± 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.

2.4.3 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.

2.5 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, 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 et al., 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 et al., 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 in the final stage of acquisition (Eckman et al., 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.

Chapter 3

Mandarin Chinese-Speaking Learners' Acquisition of Italian Consonant Length Contrast

[Feng, Q., & Busà, M. G. (2022). Mandarin Chinese-speaking learners' acquisition of Italian consonant length contrast. *System*.
https://doi.org/10.1016/j.system.2022.102938]

3.1 Introduction

Acquiring L2 consonant length contrast can be quite challenging for Mandarin Chinese-speaking learners (henceforth Chinese learners), and in fact they have been defined as “quantity-insensitive” language speakers (Tsukada et al., 2014). However, investigations on Chinese learners’ perception and/or production of L2 consonant quantity contrast have focused mainly on L2 Japanese (see §1.1 for a review), while the acquisition of consonant length contrast in L2 Italian is much less investigated. Since the actual implementation of consonant length contrast is rather language-specific and differs in Italian and Japanese to a considerable extent (Kawahara, 2015; Tanaka, 2017; Tsukada et al., 2018), by investigating the L1 Mandarin Chinese-L2 Italian language pair, this article aims to contribute to a better understanding of Chinese learners’ acquisition mechanisms of L2 consonant quantity contrast.

3.1.1 Chinese learners’ acquisition of Japanese consonant length contrast

In Japanese, length is used contrastively mainly for obstruent and nasal consonants (Kubozono, 2013). The time-related parameter (i.e., duration) is the primary acoustic cue that differentiates Japanese singleton consonants from their geminate counterparts (Kawahara, 2015). The geminate to singleton consonant duration ratio in Japanese is usually greater than 2:1 (2.8:1 in Han, 1992; 2.4:1 in Toda, 2003). The duration interplay between Japanese singleton/geminate consonants and preconsonantal vowels is the opposite to the typological tendency reported in Maddieson (1985), that is, singleton consonants are preceded by longer vowels and geminates by shorter vowels. In other words, Japanese geminate consonants are preceded by longer vowels, and singletons by shorter vowels (Han, 1994; Idemaru & Guion, 2008).

In the perception of L2 Japanese singleton/geminate consonant contrast, the identification accuracy of beginning Chinese learners seems to be affected by Japanese word pitch patterns. As shown by Minagawa and Kiritani (1996), when the pitch pattern is LH (low high), Chinese learners tend to make more errors in correctly identifying Japanese geminate consonants than singleton consonants (respectively about 75% accuracy vs. about 79% accuracy[5]). On the other hand, when the pitch pattern is HL (high low), Chinese learners' misperception direction is the opposite: it is about 88% accurate for geminate consonants and about 65% accurate for singletons. As for advanced Chinese learners of Japanese, in Masuko and Kiritani (1990), the reported perceptual identification accuracy is 78% and 77% respectively for Japanese voiceless fricative and voiceless stop consonant length contrasts. In Tsukada and Hajek (2020), advanced Chinese learners have 90% accuracy for Japanese geminate obstruent consonants and 83% accuracy for singletons.

In producing L2 Japanese consonant length contrast, beginning Chinese learners tend to produce Japanese singleton stops that are longer than one mora unit. As a result, about 35% of the consonants that are intended to be produced as singletons are misperceived as geminates by native Japanese listeners (Minagawa-Kawai & Kiritani, 1998). Intermediate level Chinese learners of Japanese consistently produce geminate consonants that are longer than singleton consonants (Lee et al., 2018; Lu et al., 2016). However, they fail to approximate the native norms in producing preconsonantal vowel duration ratios (Lu et al., 2016). That is, their pre-singleton vowels do not differ in duration from their pre-geminate vowels (Lee et al., 2018). In Yamakawa et al. (2021), Chinese learners of Japanese at a higher level than intermediate mispronounce more singleton stops as geminates than geminates as singletons. Moreover, no duration differences are found between their pre-singleton and pre-geminate syllables.

The above studies show that, in acquiring L2 Japanese consonant length contrast, Chinese learners' perceptual accuracy is always above chance level. This means that Chinese learners are able to differentiate perceptually between Japanese

[5] In Minagawa & Kiritani (1996), the authors did not provide the participants' precise accuracy rates. The data reported here were drawn from Figure 1 (p. 25).

singleton and geminate consonants to a certain extent. In production, Chinese learners generally produce geminates that are longer than singletons, which causes the majority of their consonant length production to be perceived as intended. However, in some cases, their singleton consonants sound so long that they are perceived as geminates by native Japanese listeners. As for preconsonantal vowels, Chinese learners make no difference in pre-singleton and pre-geminate vowel durations, unlike native Japanese speakers, who alternate between shorter vowels before singleton consonants and longer vowels before geminates.

3.1.2 Consonant length in Italian and Mandarin Chinese

Italian has a great number of disyllabic minimal pairs contrasting in word-internal intervocalic consonant length (e.g., *nono* 'ninth' vs. *nonno* 'grandfather', *caro* 'dear' vs. *carro* 'wagon', etc.). These consonants can be stops (/b, d, g, p, t, k/), affricates (/tʃ, dʒ/), fricatives (/f, v, s/), nasals (/m, n/) and liquids (/l, r/) (for stops, see Esposito & Di Benedetto, 1999; for affricates and fricatives, see Di Benedetto & De Nardis, 2021a; for nasals and liquids, see Di Benedetto & De Nardis, 2021b). In all cases, geminate consonants are longer than their singleton counterparts. At a regular speaking rate, geminate to singleton duration ratios vary from less than 2:1 (in affricates and fricatives) to greater than 2:1 (in stops, nasals and liquids).

Italian geminate consonants are usually considered ambisyllabic (Payne, 2005). That is, the two consonants are treated respectively as the coda of the preceding syllable and the beginning of the following one. Therefore, in Italian disyllabic minimal pairs contrasting in consonant length, the first syllables are closed in geminate words and open in singleton words. Since in Italian vowels are short in closed syllables and long in open syllables (Kramer, 2009), in disyllabic minimal pairs contrasting in singleton/geminate consonants, pre-geminate vowels are always shorter than the corresponding pre-singleton ones. The reduction ratios vary from –25% (in affricates) to –41% (in liquids).

On the other hand, in Mandarin Chinese there is no consonant length distinction at the phonemic level. Among all Mandarin Chinese consonants, only /n/ and /ŋ/ can occur in syllabic coda position; while /ŋ/ cannot occur in onset position (Duanmu, 2007). Therefore, when a syllable with nasal coda /n/ is followed by another syllable with nasal onset /n/ (e.g., <半年> *bannian*, 'half year', /pannjɛn/), together they form a "fake" or "derived" geminate consonant (i.e., a sequence of two identical consonants that do not occur morpheme-internally but in the concatenation of morphemes or words; Kubozono, 2017). However, it should be noted that "the potential for fake geminates in Mandarin is very restricted as it only occurs through concatenation of words with identical nasals across syllables in phrases or disyllabic compounds" (Meng et al., 2021, p. 3). In addition, in Mandarin Chinese, vowel length is never contrastive but only a by-product of syllable type. Specifically, vowels are long in open syllables and short in closed and weak syllables (Duanmu, 2007). Therefore, following Tsukada et al. (2014), Mandarin Chinese speakers can be considered "quantity-insensitive".

3.1.3 Theoretical framework

Among currently influential L2 speech acquisition theories, the two that have included duration properties in predicting L2 speech acquisition are the Second Language Linguistic Perception model (L2LP) proposed by Escudero (2005) and the Feature Model proposed by Brown (1998). According to the L2LP model, which focuses on L2 sound perception, at the initial stage of L2 learning, the learners tend to duplicate their L1 perception grammar in their L2 perception. Yet, through increased exposure to the L2 perceptual input, L2 learners modify their L2 perception grammar and ultimately become native-like. In our case, since in Mandarin Chinese phonology only singleton consonants exist, we speculate that Chinese learners with little Italian learning experience may tend to categorize both Italian singleton and geminate consonants as singletons ("a new scenario" according to the L2LP model) and therefore make more errors in identifying Italian geminate consonants than in identifying singletons. However, with an increased learning experience, Chinese learners may gradually approximate the L2 native norms in perceptually differentiating between Italian consonant length contrast.

According to the Feature Model, the features exploited/absent in L1 phonology can either facilitate or hinder learners' acquisition of L2 segmental contrast, depending on the availability of redeployment in L2 pronunciation learning. Specifically, in the case of L2 consonant length contrast acquisition, if the feature [± long] is exploited in the learners' native language, it will be transferred and reused in L2 learning and thereby will favor learners' perception and production accuracy. On the contrary, learners that are not familiar with the feature [± long] through their L1 phonology will not have this inherent advantage. Since Chinese learners are "quantity-insensitive", it is highly possible that they may have difficulty producing Italian singleton/geminate consonant contrast properly. However, the Feature Model does not take L2 learners' learning experience into account. Since L2 production is usually believed to be closely related to perception (SLM; Flege, 1995), we expect that Chinese learners' improved perception as a result of their increased learning experience will lead to more native-like production.

3.1.4 Research questions and hypotheses

Based on the above speculations, the research questions (RQ) and the corresponding hypotheses (H) are formulated as follows:

RQ1. How do Chinese learners perceive Italian consonant length contrast?

Our hypotheses are that Chinese learners with little Italian learning experience will tend to misidentify perceptually more geminates as singletons than singletons as geminates (H1); and more experienced learners will outperform less experienced ones in perceptual identification accuracy (H2).

RQ2. How do Chinese learners produce Italian singleton/geminate consonants and preconsonantal vowels?

Our hypothesis is that Chinese learners may not produce Italian consonant length contrast in a native-like manner (H3). However, the more experienced the learners become, the more they will be able to approximate native Italian speakers' values in producing Italian geminate to singleton consonant duration ratio and pre-singleton to pre-geminate vowel duration ratio. In other words, more experienced learners will be closer to the L2 native norms than less experienced ones (H4).

RQ3. What role does learning experience play in Chinese learners' acquisition of Italian consonant length contrast?

We hypothesize that an increased learning experience will play a positive role in enhancing Chinese learners' acquisition of Italian singleton/geminate consonant contrast (H5).

To verify the hypotheses, a perception experiment and a production experiment were carried out.

3.2 Perception experiment

3.2.1 Method

a. Participants

The perception experiment had 30 Chinese learners and 10 native Italian speakers as participants. The 30 Chinese learners were undergraduate students majoring in Italian at Dalian University of Foreign Languages in China. At the time of the experiment, none of them had had a study abroad experience in Italy. They were subdivided according to their Italian learning experience as follows: 10 first-year students (1MC group: Female = 9, Male = 1, Mean age = 19.1), 10 second-year students (2MC group:

Female = 10, Male = 0, Mean age = 19.9), and 10 third-year students (3MC group: Female = 8, Male = 2, Mean age = 20.9). The relatively restricted number of participants was due to the various restrictions imposed by the Covid-19 pandemic. The Chinese learners' gender distribution reflects the imbalance of the students' enrollment in the degree course. According to the Chinese learners' self-reports, they had received little to no specific training on the pronunciation of L2 Italian consonant length contrast at the time of the experiment. The 10 native Italian speakers (NIT group: Female = 7, Male = 3, Mean age = 21.8) were undergraduate students from the Veneto region in the North-East of Italy. Though northern Italian speakers are conventionally believed to tend to degeminate consonants (e.g., Canepari & Giovannelli, 2008; Cavanaugh, 2005), Mairano and De Iacovo (2020) show that this is not the case. No participant reported any hearing impairment at the time of the experiment.

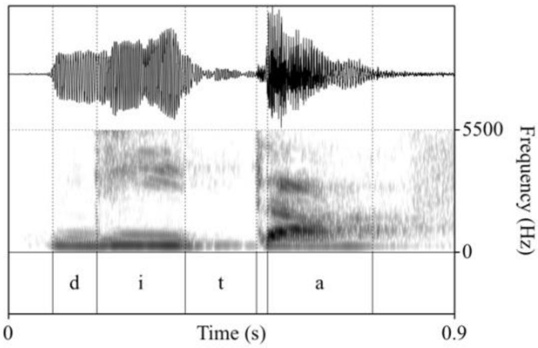
b. Materials

The target stimuli of the perception experiment were ten Italian disyllabic minimal pairs contrasting in consonant length (see Appendix II). The target consonants consisted of all types of consonants that can be geminated, namely stop (/p, t, k, d/), affricate (/tʃ, dʒ/), fricative (/f, v/), nasal (/n/) and liquid (/l/) consonants. Besides, an equal number of disyllabic minimal pairs contrasting in stop voicing (i.e., voiced vs. voiceless) served as distractors. The carrier sentence was *Leggo* ___ *bene* ('I read ___ well').

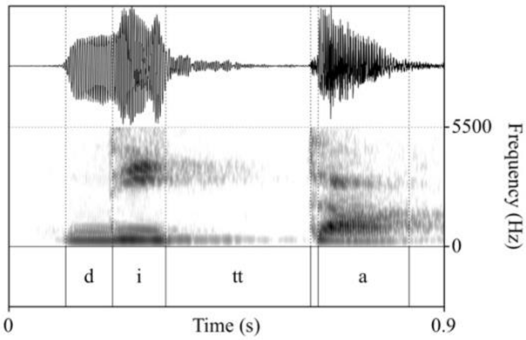
Since it has been shown that female speakers tend to adhere more consistently to standard pronunciations than male speakers (Adda-Decker & Lamel, 2005), a female native Italian speaker (age = 24) from the Veneto region was recruited for the recording session. Only one speaker's recordings were used because we did not want to create a high-variability testing condition by using multiple speakers, as the primary purpose of the present experiment was to examine the Chinese learners' perceptual accuracy of Italian consonant length contrast, rather than their ability to normalize for speakers' variability. The female native Italian speaker was instructed to read the randomly arranged word stimuli first in isolation then in the carrier sentences at a normal speech rate. The recording took place in the Language and Communication 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. A total of 80 trials ([10 target minimal pairs + 10 distracting minimal pairs] × 2 words × 2 contexts = 80) were produced. The auditory intelligibility of these trials was perceptually verified by the second author, who is a native Italian-speaking phonetician. See Table 8 for the duration information of these trials. See also Figure 5 for the acoustic waveforms and spectrograms of some of these trials. Afterwards, the trials produced were segmented into isolated words and sentences and saved as separate audio files. The intensity of the audio files was normalized to 70 dB.

Table 8. Duration information of the trials produced by the female native Italian speaker in the recording session: mean consonant durations (in ms; SDs in parentheses), mean geminate to singleton consonant duration ratios (G:S ratio; SDs in parentheses), mean preconsonantal vowel durations (in ms; SDs in parentheses) and mean pre-singleton to pre-geminate vowel duration ratios (PS:PG ratio; SDs in parentheses).

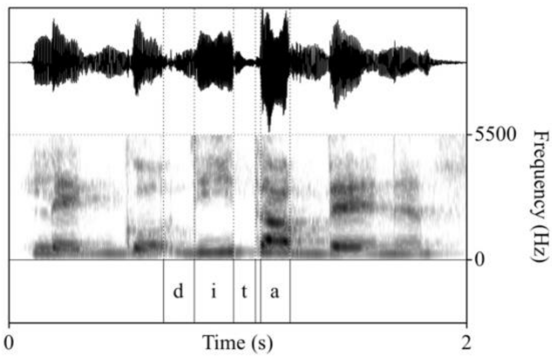
	In isolation			In carrier sentence		
Consonant	Singleton	Geminate	G:S ratio	Singleton	Geminate	G:S ratio
	100.62 (26.8)	246.78 (38.2)	2.55 (0.5)	89.48 (24.6)	223.26 (33.3)	2.58 (0.4)
Preconsonantal vowel	Pre-singleton	Pre-geminate	PS:PG ratio	Pre-singleton	Pre-geminate	PS:PG ratio
	219.85 (21.1)	121.97 (27.7)	1.91 (0.6)	178.13 (24.5)	112.80 (25.3)	1.62 (0.3)



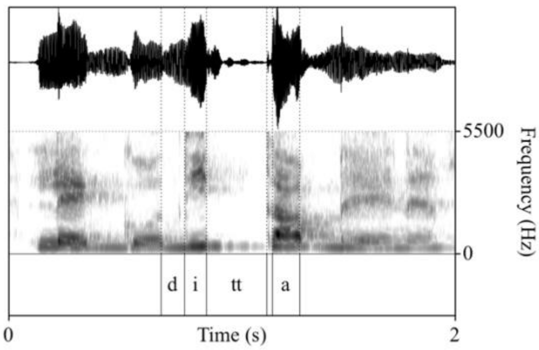
dita



ditta



Leggo dita bene.



Leggo ditta bene.

Figure 5. Acoustic waveforms and spectrograms at a 90 ms time scale of (i) *dita* 'fingers' and (ii) *ditta* 'company' in isolation (first and second panel, respectively), and at a 200 ms time scale of (iii) *dita* 'fingers' and (iv) *ditta* 'company' in the carrier sentence *Leggo* ___ *bene* 'I read ___ well' (third and fourth panel, respectively) produced by the female native Italian speaker in the recording session.

c. *Procedure*

The audio files were uploaded to the Alchemer platform (www.alchemer.com/) to create an identification test, which consisted of two parts. In the first part, the 40 isolated words were ordered randomly. For each word, two options (i.e., a minimal pair) were given on the screen. The participants, seated in front of a computer with headphones on, had to click on the option corresponding to the word heard. Each word was repeated only once. The participants were asked to listen to each word for a maximum of three times. The second part of the experiment was structured similarly to the first one except that the stimuli were the 40 sentences. The Chinese learners took the identification test in a multimedia speech lab at Dalian University of Foreign Languages in China. The native Italian speakers took the test in the Language and Communication Lab of the University of Padova in Italy.

d. *Analyses*

In total, we obtained 1600 responses for the target trials (10 target minimal pairs × 2 words × 2 contexts × 40 participants = 1600). A generalized linear mixed-effects model (GLMM) with a binomial link function was applied to the participants' responses (i.e., correct vs. incorrect) using the lme4 package 1.1.26 (Bates et al., 2015) in R 3.6.3 (R Core Team, 2020), with Group (four levels: 1MC, 2MC, 3MC, NIT), Consonant Type (two levels: Singleton and Geminate), Context (two levels: In isolation and In carrier sentence) and their interactions as fixed factors, and Subject and Item as random intercepts. The coding scheme used was treatment coding. The assessment of the main effects of the fixed factors was performed with the Type II

Wald chi-squared tests using the car package 3.0.10 (Fox and Weisberg, 2019). Post-hoc pairwise comparisons with FDR (false discovery rate) correction were carried out using the emmeans package 1.5.3 (Lenth, 2020).

3.2.2 Results

Figure 6 illustrates the four groups' perceptual accuracy rates calculated by Consonant Type and Context. As can be seen, both in isolation and in the carrier sentences, the NIT group's accuracy rates reached the ceiling for both the singleton and geminate consonants. On the other hand, in both contexts, the 1MC, 2MC and 3MC groups' accuracy rates for both types of consonants were all above chance level (hovered between 70% and 81%), but always lower than those of the NIT group. Moreover, the three learner groups' accuracy rates were rather close to each other.

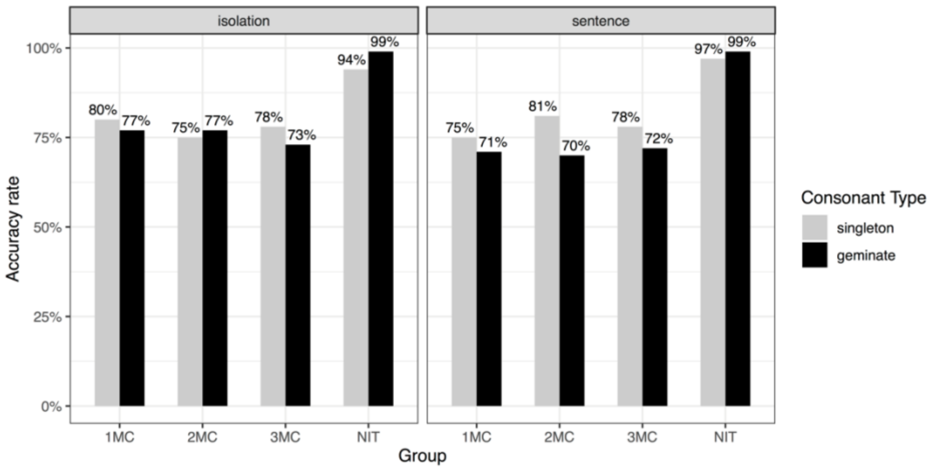


Figure 6. The 1MC, 2MC, 3MC and NIT groups' perceptual accuracy rates for the singleton and geminate consonants in isolation and in the carrier sentences.

The results of the GLMM applied to the four groups' responses are shown in Table 9. As can be seen, the GLMM yielded a significant main effect on Group, indicating that the participants' perceptual accuracy differed significantly by group. Post-hoc pairwise comparisons revealed that there were no significant differences across the three learner groups (1MC vs. 2MC: $\beta = 0.005$, $SE = 0.32$, $z = 0.02$, $p =$

0.99; 1MC vs. 3MC: $\beta = -0.06$, $SE = 0.32$, $z = -0.18$, $p = 0.99$; 2MC vs. 3MC: $\beta = -0.06$, $SE = 0.32$, $z = -0.19$, $p = 0.99$), which were all significantly different from the NIT group (1MC vs. NIT: $\beta = 2.82$, $SE = 0.50$, $z = 5.67$, $p < .0001$; 2MC vs. NIT: $\beta = 2.82$, $SE = 0.50$, $z = 5.66$, $p < .0001$; 3MC vs. NIT: $\beta = 2.88$, $SE = 0.50$, $z = 5.78$, $p = p < .0001$). These results indicate that the Chinese learners with different learning experiences identified perceptually Italian consonant length contrast with an equal level of accuracy; and they were always significantly less accurate than the native Italian listeners. Also, it should be noted that there was no significant main effect on Consonant Type, indicating that the participants' identification accuracy did not vary by this factor. Moreover, the nonsignificant interaction of Group × Consonant Type × Context indicates that the ways in which the perceptual identification accuracy differed between groups did not vary by consonant type and context.

Table 9. Results of the GLMM of the participants' perceptual accuracy of Italian consonant length contrast (*p* values < 0.05 are in bold).

	Fixed effects			Random effects	
				By Subject	By item
	χ^2	*df*	*p*	*SD*	*SD*
Intercept	-	-	-	0.36	0.15
Group	31.95	3	**< .001**	-	-
Consonant Type	0.72	1	0.40	-	-
Context	0.48	1	0.49	-	-
Croup × Consonant Type	5.32	3	0.15	-	-

Croup × Context	1.92	3	0.59	-	-
Consonant Type × Context	1.19	1	0.27	-	-
Croup × Consonant Type × Context	1.62	3	0.66	-	-

3.2.3 Discussion

As expected, compared to native Italian listeners, Chinese learners are less accurate in perceptually identifying Italian singleton and geminate consonants. However, some results are unexpected.

First, though Chinese learners are "quantity-insensitive" language speakers, in the majority of cases (more than 70% of instances), they succeed in correctly differentiating between Italian singleton and geminate consonants. We argue that though speakers' perceptual sensitivity to duration may be reduced if length distinctions are absent in their L1 (Gottfried & Beddor, 1988), it does not mean that in processing linguistic stimuli, the so-called "quantity-insensitive" language speakers completely lose the perceptual sensitivity to segment duration differences. In fact, "perceptual sensitivity to duration is needed in all languages as segment durations do not only signal lexical contrasts but may also convey information about prosodic structure" (Altmann et al., 2012, p. 389). For example, it has been noted that neutral-tone syllables in Mandarin Chinese are shorter than lexical-tone syllables (Lee & Zee, 2008; Lin & Yan, 1980), and duration is the primary cue by which Mandarin Chinese listeners differentiate the two types of syllables (Lin, 1985). Thus, Mandarin Chinese speakers may not be completely "quantity-insensitive" as is commonly believed. Moreover, similar to what happens with the singleton/geminate consonant contrast in Kelantan Malay (Hamzah, 2010), Swedish and Arabic (Hassan, 2002), the duration differences between Italian singleton consonants and their geminate counterparts are over the threshold of the Just Noticeable Difference (JND) for segment duration

(Payne, 2005; see Klatt [1976] and Lehiste [1970] for JND), which means that the duration differences existing in Italian consonant length contrast are in most cases perceptually discernible. Therefore, though Chinese learners are not as sensitive as native Italian listeners to Italian consonant duration differences - because of the influence of their L1 Mandarin Chinese phonology, they are able to distinguish perceptually between Italian singleton and geminate consonants to a certain extent.

Second, though geminate but not singleton consonants are absent in Mandarin Chinese, contrary to our H1, Chinese learners do not misperceive geminates more than singletons. We argue that this is because in the present study, even the Chinese learners with the least learning experience had studied Italian for one year at the time of the experiment. Therefore, it may not be appropriate to view them as beginning learners. In other words, they might have already passed the initial L2 learning stage as defined by the L2LP model (Escudero, 2005), that is, the stage at which L2 learners tend to duplicate their L1 perception grammar in their L2 perception, and have established two consonant length categories (i.e., short and long) to address the new duration distribution in L2 Italian. This fact may imply that Chinese learners' new duration category does not take a long time to be formed.

Third, contrary to our H2, Chinese learners' increased learning experience does not result in higher accuracy rates. This fact raises the question of what role Chinese learners' learning experience plays in their L2 quantity acquisition, which will be further discussed later.

To test how Chinese learners' perception of Italian consonant length contrast maps onto their production pattern, a production experiment was carried out.

3.3 Production experiment

3.3.1 Method

a. Participants

The participants of the production experiment were the same as those in the perception experiment. None of them reported any speech impairment at the time of the experiment.

b. Materials

To maintain consistency between the production and perception experiments as well as prevent unfamiliar sounds from affecting the Chinese learners' production, five minimal pairs that served as target stimuli in the perception experiment and had target consonants available in both Italian and Mandarin Chinese phonology (i.e., /p, t, k, n, l/) were selected as target stimuli for the production experiment (see Appendix II). The distractors were the five minimal pairs contrasting in stop voicing that were used in the perception experiment. The carrier sentence was still *Leggo* ___ *bene* ('I read ___ well'). The target word stimuli were first repeated twice randomly in isolation. Then they were inserted in the carrier sentences and repeated twice in random order. The stimuli in isolation and in the carrier sentences were printed on a paper sheet for the subjects to read. After every 20 words/sentences, a cartoon picture was inserted to remind the participants to have a short break.

c. Procedure

In the production experiment, the participants were instructed to read the stimuli on the paper sheet at a normal speed. They were also asked to make a short break when the cartoon pictures told them to. The recordings of the native Italian speakers took place in the Language and Communication 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 Chinese learners 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 target intervocalic consonants and the preconsonantal vowels were labeled in Praat (Boersma & Weenink, 2020). As shown by Esposito and Di Benedetto (1999), closure duration is the primary cue for gemination in Italian stops. Therefore, the intervocalic stop consonants in the present experiment were labeled from the offset of the periodic wave of the preceding vowel to the release burst of the stop consonant; the VOT was not included. For the target intervocalic lateral and nasal consonants, their boundaries with the adjacent vowels were identified where abrupt spectral changes were observed. For the preconsonantal vowels, boundaries were located at the onset and offset of periodicity in the acoustic waveform. Afterwards, the durations of the target intervocalic consonants and the preconsonantal vowels were extracted using a Praat script (Lennes, 2002).

e. Analyses

The duration values were converted into duration ratios for statistical analyses. A total of 1600 target tokens were elicited (10 target word stimuli × 2 contexts × 2 repetitions × 40 participants = 1600). Of these tokens, half had intervocalic singleton consonants, and the other half had intervocalic geminate consonants. Therefore, we should have had 800 geminate to singleton consonant duration ratios and 800 pre-singleton to pre-geminate vowel duration ratios. However, in measuring duration values, 56 unmeasurable/misread tokens with intervocalic consonants (26 singletons, 30 geminates) and 61 unmeasurable/misread tokens with preconsonantal vowels (28 pre-singleton and 33 pre-geminate) were discarded. Because of these missing values, in converting duration values into duration ratios, we obtained 751 effective consonant duration ratios and 747 effective preconsonantal vowel duration ratios.

Two linear mixed models (LMM) were applied respectively to the consonant and preconsonantal vowel duration ratios using the lme4 package 1.1.26 (Bates et al., 2015) in R 3.6.3 (R Core Team, 2020), with Group (four levels: 1MC, 2MC, 3MC, NIT), Context (two levels: In isolation and In carrier sentence) and their interaction as fixed factors, and Subject and Item as random intercepts. The coding scheme used was treatment coding. 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 & Weisberg, 2019). Post-hoc pairwise comparisons with FDR (false discovery rate) correction were carried out using the emmeans package 1.5.3 (Lenth, 2020).

3.3.2 Results

a. Consonant duration ratios

Table 10 lists the mean duration values of the consonants produced by the 1MC, 2MC, 3MC and NIT groups and their mean geminate to singleton consonant duration ratios. Figure 7 shows a visual representation of the distribution of the four groups' consonant duration ratios. As Table 10 and Figure 7 show, both in isolation and in the carrier sentences, all four groups had longer durations for geminate than for singleton consonants. However, the NIT group had greater consonant duration ratios than the three learner groups. Besides, in terms of duration value, the singleton consonants produced by the 1MC, 2MC, 3MC groups were consistently longer than those of the NIT group; while the four groups' geminate consonants had durations that did not differ to a great extent both in isolation and in the carrier sentences.

Table 10. Mean consonant durations (in ms; SDs in parentheses) and mean geminate to singleton consonant duration ratios (G:S ratio; SDs in parentheses) of the 1MC, 2MC, 3MC and NIT groups.

	In isolation			In carrier sentence		
	Singleton	Geminate	G:S ratio	Singleton	Geminate	G:S ratio
1MC	131.2	171.6	1.37	100.3	122.7	1.32

	(41.7)	(66.9)	(0.53)	(55.3)	(64.1)	(0.68)
2MC	143.4 (45.0)	182.3 (58.7)	1.31 (0.35)	105.0 (39.6)	128.2 (53.2)	1.29 (0.52)
3MC	140.0 (75.8)	169.5 (78.0)	1.37 (0.83)	106.7 (46.8)	134.3 (71.5)	1.38 (0.72)
NIT	80.5 (28.0)	183.5 (43.2)	2.61 (1.48)	69.6 (22.4)	133.3 (33.0)	2.11 (0.93)

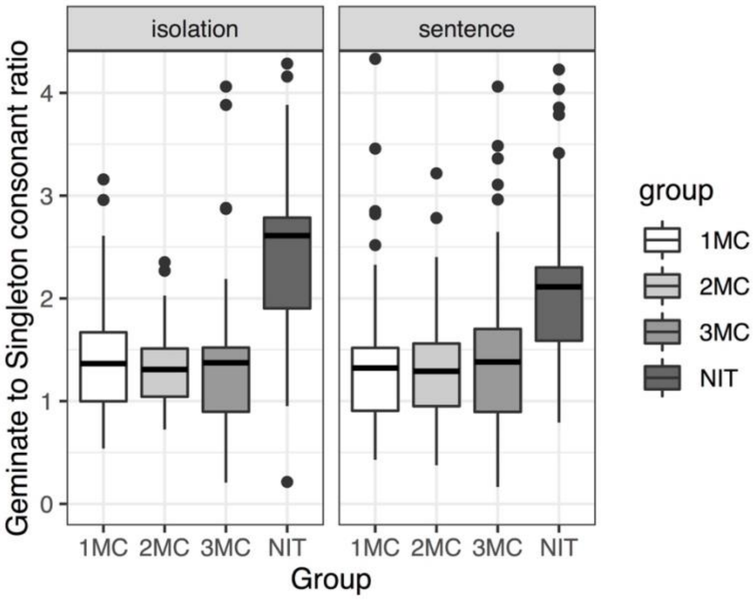


Figure 7. Distribution of the geminate to singleton consonant duration ratios of the 1MC, 2MC, 3MC and NIT groups.

For the statistical analyses, the consonant duration ratios were normalized using the bestNormalize package 1.7.0 (Peterson & Cavanaugh, 2020). After fitting the LMM, the visual diagnostics of the histogram and the plot of residuals showed no drastic violations of the assumptions of normality and homoscedasticity.

Table 11 shows the results of the LMM. As can be seen, the LMM yielded significant main effects on Group and Context, indicating that the participants' geminate to singleton consonant duration ratios differed significantly by these two factors. Post-hoc pairwise comparisons revealed that, first, all three learner groups were significantly different from the NIT group (1MC vs. NIT: $\beta = -1.17$, $SE = 0.17$, $t(36.4) = -7.02$, $p < .0001$; 2MC vs. NIT: $\beta = -1.14$, $SE = 0.17$, $t(35.3) = -6.94$, $p < .0001$; 3MC vs. NIT: $\beta = -1.18$, $SE = 0.17$, $t(35.2) = -7.19$, $p = p < .0001$), indicating that though the Chinese learners were able to produce short-long differences for Italian consonant length contrast, their geminate to singleton consonant duration ratios were significantly smaller than those of the native Italian speakers. Second, there were no significant differences across the learner groups (1MC vs. 2MC: $\beta = -0.02$, $SE = 0.17$, $t(36.9) = -0.13$, $p = 0.92$; 1MC vs. 3MC: $\beta = 0.02$, $SE = 0.17$, $t(36.8) = 0.11$, $p = 0.92$; 2MC vs. 3MC: $\beta = 0.04$, $SE = 0.17$, $t(35.7) = 0.24$, $p = 0.92$), indicating that though the Chinese learners had different learning experiences, they produced Italian geminate to singleton consonant duration ratios very similarly. Third, the participants performed significantly differently in the two reading contexts (in isolation vs. in the carrier sentences: $\beta = 0.15$, $SE = 0.06$, $t(704) = 2.643$, $p = 0.0084$). That is, the participants had larger consonant duration ratios for the words produced in isolation than for those produced in the carrier sentences. Also, there was no significant interaction between Group and Context, indicating that the differences in consonant duration ratio between groups did not vary by context.

Table 11. Results of the LMM of the participants' production of geminate to singleton consonant duration ratios (*p* values < 0.05 are in bold).

	Fixed effects			Random effects	
				By Subject	By item
	χ^2	*df*	*p*	*SD*	*SD*
Intercept	-	-	-	0.10	0.01

Group	74.88	3	**< .001**	-	-
Context	7.32	1	**0.0068**	-	-
Croup × Context	6.58	3	0.086	-	-

b. Preconsonantal vowel duration ratios

The mean duration values of the preconsonantal vowels produced by the 1MC, 2MC, 3MC and NIT groups and their mean pre-singleton to pre-geminate vowel duration ratios are shown in Table 12. Figure 8 shows a visual representation of the distribution of the four groups' preconsonantal vowel duration ratios. As Table 12 and Figure 8 show, only the NIT group had pre-singleton to pre-geminate vowel duration ratios greater than 1, both in isolation and in the carrier sentences. The preconsonantal vowel duration ratios of the three learner groups hovered around 1. This means that only the NIT group produced pre-singleton vowels that were longer than pre-geminate vowels. The three learner groups, on the other hand, produced pre-singleton and pre-geminate vowels with similar durations.

Table 12. Mean preconsonantal vowel durations (in ms; SDs in parentheses) and mean pre-singleton to pre-geminate vowel duration ratios (PS:PG ratio; SDs in parentheses) of the 1MC, 2MC, 3MC and NIT groups.

	In isolation			In carrier sentence		
	pre-singleton	pre-geminate	PS:PG ratio	pre-singleton	pre-geminate	PS:PG ratio
1MC	189.8 (51.9)	198.9 (52.7)	0.97 (0.24)	154.4 (40.7)	161.1 (39.0)	0.97 (0.23)
2MC	167.8	176.8	1.00	155.0	167.7	0.97

	(52.9)	(62.7)	(0.23)	(48.3)	(60.0)	(0.21)
3MC	147.0	152.5	0.99	145.4	151.5	1.00
	(38.4)	(39.5)	(0.21)	(41.1)	(46.5)	(0.29)
NIT	149.5	103.7	1.48	119.3	90.0	1.32
	(37.1)	(28.2)	(0.33)	(39.2)	(19.3)	(0.31)

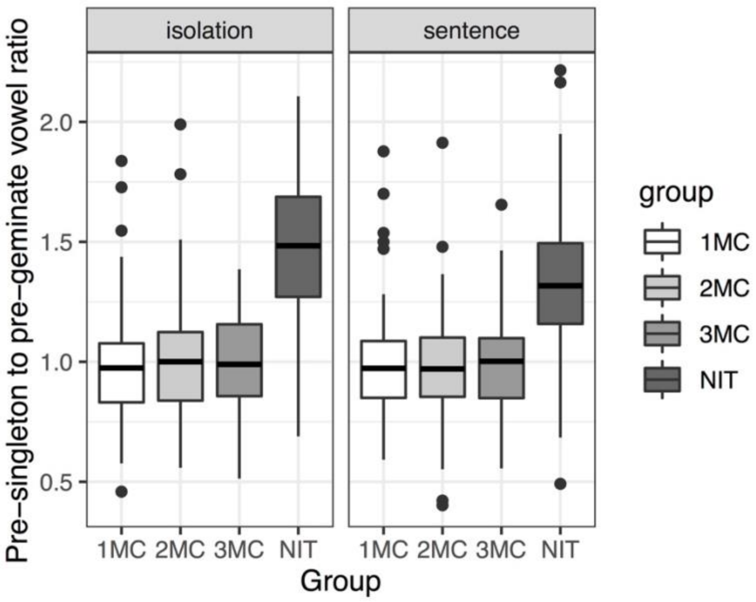


Figure 8. Distribution of the pre-singleton to pre-geminate vowel duration ratios of the 1MC, 2MC, 3MC and NIT groups.

For the statistical analyses, the preconsonantal vowel duration ratios were normalized using the bestNormalize package 1.7.0 (Peterson & Cavanaugh, 2020). After fitting the LMM, the visual diagnostics of the histogram and the plot of residuals showed no drastic violations of the assumptions of normality and homoscedasticity.

Table 13 shows the results of the LMM, which yielded significant main effects on all fix factors, namely Group, Context, and their interaction. Table 14 summarizes the results of the post-hoc pairwise comparisons. As can be seen, the three learner groups were always significantly different from the NIT group, indicating that while the native Italian speakers alternated between short vowels before geminate consonants and long vowels before singletons, the Chinese learners produced no difference in pre-singleton and pre-geminate vowel durations. Moreover, no significant between-group differences were found across the 1MC, 2MC and 3MC groups, indicating that the Chinese learners' different learning experiences did not affect their productions of Italian preconsonantal vowel duration ratios. Also, the within-group analysis showed that only the NIT group performed differently in the two reading contexts (i.e., in isolation vs. in the carrier sentences). That is, the duration differences between the pre-singleton and pre-geminate vowels produced by the native Italian speakers were more evident when the words were produced in isolation than in the carrier sentences.

Table 13. Results of the LMM of the participants' production of pre-singleton to pre-geminate vowel duration ratios (*p* values < 0.05 are in bold).

	Fixed effects			Random effects	
				By Subject	By item
	χ^2	*df*	*p*	*SD*	*SD*
Intercept	-	-	-	0.06	0.01
Group	116.22	3	**< .001**	-	-
Context	6.14	1	**0.013**	-	-
Croup × Context	9.88	3	**0.020**	-	-

Table 14. Results of the post-hoc pairwise comparisons of the pre-singleton to pre-geminate vowel duration ratios of the 1MC, 2MC, 3MC and NIT groups (*p* values < 0.05 are in bold).

	Estimate	*SE*	*df*	*t* ratio	*p* value

In isolation					
1MC vs. 2MC	–0.10	0.17	68.8	–0.61	0.81
1MC vs. 3MC	–0.06	0.17	70.1	–0.37	0.81
2MC vs. 3MC	0.04	0.16	64.2	0.25	0.81
1MC vs. NIT	–1.48	0.17	67.3	–9.00	**<.0001**
2MC vs. NIT	–1.38	0.16	61.6	–8.59	**<.0001**
3MC vs. NIT	–1.42	0.16	62.6	–8.80	**<.0001**
In carrier sentence					
1MC vs. 2MC	0.004	0.16	65.7	0.03	0.98
1MC vs. 3MC	–0.07	0.16	64.4	–0.46	0.78
2MC vs. 3MC	–0.08	0.16	62.5	–0.49	0.78
1MC vs. NIT	–1.05	0.16	64.4	–6.45	**<.0001**
2MC vs. NIT	–1.05	0.16	62.5	–6.53	**<.0001**
3MC vs. NIT	–0.98	0.16	61.2	–6.08	**<.0001**
In isolation vs. In carrier sentence					
1MC	0.006	0.12	702.9	0.05	0.96
2MC	0.11	0.12	700.2	0.96	0.34
3MC	–0.007	0.12	700.4	–0.06	0.95
NIT	0.44	0.11	699.6	3.88	**0.0001**

3.3.3 Discussion

The productions of the native Italian speakers in the present experiment show that northern Italian speakers consistently and properly produce Italian geminate consonants - in contrast to the conventional claim that northern Italian speakers tend to degeminate consonants (Canepari & Giovannelli, 2008; Cavanaugh, 2005). This confirms what was found in Mairano and De Iacovo (2020) that attribute this fact to the progressive standardization of Italian in Italy.

Concerning Chinese learners' production of Italian consonant length contrast, the results show that Chinese learners succeed in making a distinction between shorter singleton and longer geminate consonants in Italian, even when they have only one year of Italian learning experience. This parallels what was found in our perception experiment, suggesting that Chinese learners manage to form new consonant duration categories in a short time and successfully apply the perceived duration differences to their productions. Besides, it is conceivable that the orthographic difference in Italian consonant quantity contrast (i.e., one letter for singleton consonants and two letters for geminates) may also contribute to a certain extent to guiding Chinese learners to produce Italian geminate consonants that are longer than their singleton counterparts. That the orthographic forms of singleton vs. geminate consonants may influence speakers' productions has been shown, for example, by Bassetti et al. (2018), who found that L1 Italian speakers, following English orthography, produce consonant length distinctions in their L2 English speech (though with smaller duration differences than in L1 Italian).

However, Chinese learners produce significantly smaller geminate to singleton consonant duration ratios than native Italian speakers. This shows that Chinese learners fail to approximate the L2 norms in producing Italian consonant length contrast, which is in line with our H3. Nevertheless, it should be noted that Chinese learners have smaller consonant duration ratios not because they produce shorter geminate consonants, but because they produce much longer singletons as compared to native Italian speakers. Since Italian consonant length is inversely related to speaking rates (Pickett et al., 1999), we argue that it is Chinese learners' slower speaking rates that lengthen both their singleton and geminate consonants. Thus, Chinese learners approach native Italian speakers' duration values in the production of geminate but not singleton consonants. This triggers the following questions: Will a certain number of these longer singletons be perceived as geminates by native Italian listeners as is reported for L2 Japanese (Minagawa-Kawai & Kiritani, 1998; Yamakawa et al., 2021)? Will these lengthened geminate consonants be perceived as "true" geminates by native Italian listeners? To answer these questions, a future perception experiment that involves native Italian listeners as raters is needed.

As for Chinese learners' production of preconsonantal vowels in Italian consonant length contrast, our data show that Chinese learners produce no duration differences between pre-singleton and pre-geminate vowels, unlike native Italian speakers that have shorter vowels in pre-geminate position. Given that in both Italian and Mandarin Chinese vowels are short in closed syllables, we hypothesize that this is due to the fact that Chinese learners do not treat the first syllables in Italian disyllabic geminate words as closed syllables. This is because, with the exception of /n/, none of the consonants that can be geminated in Italian can occur in syllable coda position in Mandarin Chinese. Thus, influenced by the L1 norms, Chinese learners fail to consider Italian geminate consonants ambisyllabic as native Italian speakers do, but treat them simply as a whole that acts as the initial part of the subsequent syllable. In other words, Chinese learners treat the first syllables of both Italian disyllabic geminate and singleton words as open syllables, and therefore make no difference in preconsonantal vowel durations.

Regarding the effects of learning experience, our data show that, contrary to our H4, an increase in learning experience brings no improvement in Chinese learners' production of Italian consonant length contrast, which parallels what was found in our perception experiment. We will further discuss this point in § 4.

3.4 General discussion and conclusions

This study set out to investigate the acquisition of Italian consonant length contrast by Mandarin Chinese-speaking learners. Regarding the first two research questions that respectively address how Chinese learners perceive Italian consonant length contrast and how they produce Italian singleton/geminate consonants and preconsonantal vowels, our results show that, in both perception and production, Chinese learners can distinguish between the two consonant length categories to a certain extent, but not in a native-like manner. Moreover, the duration interplay between Italian consonants and preconsonantal vowels (i.e., longer vowels before singleton consonants and shorter vowels before geminate consonants) is entirely ignored by Chinese learners.

In relation to the third research question that addresses the role played by the learning experience in Chinese learners' acquisition of Italian consonant quantity distinction, the results are contrary to our H5. That is, Chinese learners' increased learning experience does not seem to enhance either their perception or production of Italian consonant quantity contrast. Previous studies focusing on L2 vowel/consonant quantity contrast acquisition have reported similar results (e.g., for perception, see Hayes, 2002; for production, see Harada, 2006; Hirata, 2017; Lee & Mok, 2018; Toda, 1997). Why is this? Some plausible reasons can be found in the literature.

Luo et al. (2020) argue that the majority of L2 quantity contrast acquisition may occur at the initial stage of L2 acquisition, of which the cut-off point should be 6-12 months according to Best & Tyler (2007). Possibly due to this, in L2 quantity contrast acquisition, learners with language learning experience always outperform those without experience (e.g., Hayes-Harb & Masuda, 2008; Kabak et al., 2011; Tsukada & Hajek, 2019a). However, once the initial stage is passed and the quantity categories are established, for most learners an increased learning experience benefits no longer L2 quantity contrast acquisition. Alternatively, considering that learners with little learning experience can achieve accuracy rates far above chance level in perception, and can make short-long differences in production, it is possible that they do *not* encounter intelligibility issues when communicating with native speakers. Thus, the urgency to further improve the L2 quantity contrast pronunciation becomes low, and the inaccuracy problems persist even when their learning experience increases (Hirata, 2017; Lee & Mok, 2018).

As an alternative explanation, we argue that Chinese learners' scarce improvement in L2 Italian consonant length contrast acquisition is caused by their insufficient L2 perceptual input. The L2LP model (Escudero, 2005) indicates that L2 learners modify their L2 perception grammar through increased exposure to the L2 perceptual input. The Chinese learners in this study were learning L2 Italian exclusively in China. Therefore, their daily-life communication is predominantly in Mandarin Chinese; and their authentic L2 Italian perceptual input is limited to lessons given by native Italian-speaking teachers that last only a few hours per week. Thus, the increase in the present Chinese learners' learning experience is not paralleled by a substantial increase in their exposure to L2 perceptual input, causing them to fail to get closer to a native level of proficiency in perception. So, the Chinese learners' scarce improvement in perception causes, in turn, their standstill in production.

This argument seems compatible with the L1 consonant quantity contrast acquisition process. That is, the perceptual input plays a vital role in consonant length distinction acquisition also for L1 learners. For example, both Finnish and Japanese have singleton/geminate consonant contrast in their phonology, though geminate consonants occur more frequently in Finnish than in Japanese. So, in Finnish children's books, geminate consonants represent a total of 9.9% of all the consonant occurrences, while this proportion is 5.1% in Japanese children's books (Aoyama, 2000). More importantly, in Finnish and Japanese mothers' speech to children, the proportions of geminate consonants in all consonant occurrences are respectively 13.2% and 6.5% (Kunnari et al., 2001). As a result, Finnish children are exposed to geminate consonant input almost twice as frequently as Japanese children. Possibly due to this input difference, L1 Finnish children acquire consonant length distinction in production more rapidly than L1 Japanese children (Aoyama, 2000; Kunnari et al., 2001).

There is another issue worth considering. Recall that the actual implementations of consonant length contrast in Italian and Japanese differ to a considerable extent (see §1.1 and §1.2 for a comparison). Firstly, consonant length is contrastive for a wider number of phonemes in Italian than in Japanese; and secondly, the duration interplay between singleton/geminate consonants and preconsonantal vowels in Italian is the opposite to that in Japanese. However, despite these differences, Chinese learners' consonant length contrast acquisition patterns for these two languages show some similarities. Specifically, in perception, Chinese learners have accuracy rates well above chance level but significantly lower than the ceiling. In production, Chinese learners produce geminate consonants that are longer than singletons, though their singletons tend to be longer than native speakers'. Besides, Chinese learners ignore the duration interplay between geminate/singleton consonants and preconsonantal vowels, and produce pre-singleton and pre-geminate vowels with similar durations. This may indicate that Chinese learners follow a similar acquisition pattern in dealing with consonant quantity contrasts in different L2s. However, at the present stage this is a purely speculative conclusion and would need to be addressed with further investigations.

Several crucial issues are left for future research. For instance, in the present perceptual identification test, some target minimal pairs consisted of words with unbalanced lexical frequency, and it is unclear whether this imbalance has affected the participants' perception. Therefore, to consolidate the current conclusions, future perception experiments using nonwords contrasting in consonant length as stimuli may help. Besides, it has been shown that the manners of articulation of consonants (i.e., sonorant vs. obstruent; Dmitrieva, 2018), the voicing feature of obstruent consonants (Tsukada & Hajek, 2019b), the length differentials of singleton/geminate consonant contrast (Hayes, 2002), and the sonority differences between consonants and post-consonantal vowels (Motohashi-Siago & Hardison, 2009) could all affect cross-linguistic perceptual accuracy of consonant quantity contrast. Therefore, it would also be interesting to know whether/how these factors interact with learning experience to affect Chinese learners' perceptual discrimination between Italian singleton and geminate consonants.

Also, some pedagogical implications can be drawn from this study. Specifically, the Chinese learners in the present study fail to fully master the duration differences creating Italian consonant length contrast (especially the duration interplay between Italian consonants and preconsonantal vowels) and to overcome on their own the acquisition standstill mainly because (i) they are relatively "quantity-insensitive" and (ii) have limited L2 Italian perceptual input - as they have learned Italian exclusively in China. Therefore, one possible way to improve Chinese learners' acquisition of Italian consonant length contrast is to increase their L2 Italian input through creating an immersion context (e.g., study-aboard context) for them. Also, considering that computer-assisted (e.g., Hirata, 2004; Motohashi-Siago & Hardison, 2009) and gesture-aided (e.g., Li et al., 2020) teaching methods, by visualizing L2 segmental duration differences through either acoustic waveforms or hand movements, have been shown to be effective in helping "quantity-insensitive" learners to enhance their L2 durational contrast acquisition, it is highly recommended to incorporate these two methods in Chinese learners' L2 Italian pronunciation training classrooms.

Chapter 4

The Acquisition of Italian /r-l/ Contrast by L1-Chinese Learners

[Feng, Q., & Busà, M. G. (in press). The acquisition of Italian /r-l/ contrast by L1-Chinese learners. *Journal of Monolingual and Bilingual Speech*.]

4.1 Introduction

The tendency to confound Italian /r/ and /l/ in both perception and production by L1-Chinese learners (henceforth Chinese learners) has been widely documented (e.g. Cao, 2018; Costamagna, 2010; Dalla Palma, 2019; D'Annunzio, 2009; Trifone, 2014). However, this issue has not been investigated empirically, as all the studies carried out so far have been based on impressionistic observations from either teaching experiences or conversational interactions. Therefore, the present study aims to fill this gap by collecting experimental data on how Chinese learners perceive and produce the /r-l/ contrast in Italian. Before looking into it, we first look at how Chinese learners acquire the /r-l/ contrast in languages other than Italian.

4.1.1 Chinese learners' acquisition of /r-l/ contrast in languages other than Italian

Since Italian /r/ is usually realized as a trill or a tap (see the next subsection for a detailed review), we are particularly interested in the languages where the rhotic /r/ can be realized as a trill and/or a tap. Among these languages, Spanish, European Portuguese, Russian and Persian are the L2/L3s in which Chinese learners' acquisition of the /r-l/ contrast has been empirically investigated.

Spanish has two rhotics, namely the tap [ɾ] and the trill [r] that contrast in intervocalic position. The primary difference between the Spanish tap and trill is that the former is produced with one rapid contact between the tongue tip and the alveolar ridge, while the latter is produced with two or more rapid contacts (see Patience [2018] for a review). Spanish /l/ is typically an alveolar lateral approximant (Martínez-Celdrán et al., 2003). In perceiving Spanish minimal pairs contrasting in /r-l/, Chinese learners have more difficulty correctly identifying the tap-lateral contrast ([ɾ-l]) than the trill-lateral contrast ([r-l]) (Chih, 2013; Ortí Mateu, 1990). In production, Chinese learners initially tend to replace both the tap [ɾ] and the trill [r] with the alveolar lateral [l]. While the production of the tap [ɾ] can be eventually acquired by Chinese learners as their Spanish language proficiency increases, this is not the case for the trill [r]. In other words, even some proficient Chinese learners of Spanish are unable to produce the trill [r] properly, and they produce taps and a number of other segments (e.g. [l], [ɹ], [dɾ], etc.) instead (Patience, 2018).

European Portuguese has two rhotics: [ɾ] and [ʀ]. These two phonemes are contrastive in intervocalic position. [ɾ] is a tap, articulated with a rapid tongue tip movement against the alveolar ridge. [ʀ] is realized as a uvular fricative. European Portuguese /l/ is realized as an alveolar lateral [l] in onset position, and as a velarised [ɫ] in coda (see Zhou et al. [2021] for a review). In perception, Chinese learners confound the intervocalic tap [ɾ] and the alveolar lateral [l] bidirectionally (Cao, 2018; Vale, 2020; Zhou, 2021); and an increased learning experience does not result in higher perceptual discrimination accuracy (Zhou, 2021). In production, the alveolar lateral [l] in onset position presents little difficulty for Chinese learners. By contrast, the velarised [ɫ] in coda position is produced with rather low accuracy. As for the rhotics, Chinese learners acquire the uvular fricative /ʀ/ moderately well. However, they produce the tap [ɾ] with lower accuracy. When failing to produce the target tap [ɾ], Chinese learners tend to use the alveolar lateral [l] as a substitute for the onset [ɾ], and to use other repair strategies (epenthesis, deletion, segmental repair) for the coda [ɾ] (Zhou et al., 2021).

Russian consonants are characterized by the phonological opposition of palatalization and non-palatalization. The non-palatalized rhotic is realized as an alveolar trill [r] (in careful pronunciation), and its palatalized counterpart is usually realized as a tap [ɾʲ] (Yanushevskaya & Bunčić, 2015). The non-palatalized /l/ is a velarized apical dental/alveolar lateral [ɫ], and its palatalized counterpart is an apical/laminal dental/alveolar lateral [lʲ] (Kochetov, 2005). In perceptual learning, Chinese learners assimilate Russian word-initial non-palatalized [ɫ] but not [r] to Mandarin Chinese [l], and they can differentiate between Russian word-initial non-palatalized [r-ɫ] contrast to a certain degree; however, Chinese learners' discrimination accuracy does not improve as a result of their increased learning experience and improved Russian proficiency (Yang & Chen, 2019).

The actual realization of Persian rhotic consonants is under debate. To summarize, it has been argued that (voiced and voiceless) trills, taps, fricatives and approximants are the allophonic variations of Persian rhotics (see Falahati [2015] for a review). In the realization of Persian rhotics by Chinese learners, though a great between-subject variation exists, the general tendency is that approximants and taps rank as the two most frequent variants, followed by trills and fricatives. Besides, the replacement of Persian rhotics by the alveolar lateral is also observed in a small number of Chinese learners (Falahati, 2015).

To sum up, previous research has revealed that the acquisition of L2/L3 /r-l/ contrast represents a challenge for Chinese learners. In perception, Chinese learners experience difficulty differentiating between these two phonemes, especially when /r/ and /l/ are realized as a tap [ɾ] and an alveolar lateral [l], respectively (Cao, 2018; Chih, 2013; Ortí Mateu, 1990; Vale, 2020; Zhou, 2021). For other contrasts (e.g. [r-l] or [r-ɫ]), Chinese learners' perceptual accuracy is relatively higher (Chih, 2013; Ortí Mateu, 1990; Yang & Chen, 2019). Besides, the effect of Chinese learners' learning experience on their perception of the /r-l/ contrast seems to be small (Yang & Chen, 2019; Zhou, 2021). In production, /r/ tends to be difficult for Chinese learners regardless of its actual realization (Patience, 2018; Zhou et al., 2021). Moreover, the trill [r] is more difficult than the tap [ɾ], and this production difficulty does not seem to decrease when proficiency increases (Patience, 2018). By contrast, Chinese learners have little difficulty producing the alveolar lateral [l]; and when failing to produce the target /r/, they show the tendency to use [l] as a substitute (Falahati, 2015; Patience, 2018; Zhou et al., 2021). Does Chinese learners' acquisition of Italian /r-l/ contrast follow a similar pathway? To answer this question, we first look at the articulatory characteristics of the /r/ and /l/ in Standard Italian and Mandarin Chinese.

4.1.2 /r/ and /l/ in Standard Italian and Mandarin Chinese

Standard Italian /r/ is realized with different allophonic variants across different phonetic contexts. Specifically, in word-initial position, Italian /r/ is usually realized as a trill (i.e. [r]) with two to three linguo-palatal contacts (Bertinetto & Loporcaro, 2005; Celata et al., 2016). While in intervocalic position, Italian trill /r/ is often reduced to a tap (i.e. [ɾ]) with only one linguo-palatal contact (Bertinetto & Loporcaro, 2005; Kramer, 2009; Rogers & d'Arcangeli, 2004). In other positions, Italian /r/ is generally pronounced as a trill [r] in stressed syllables and as a tap [ɾ] in unstressed ones (Canepari, 1999). However, it should be noted that Italian /r/ is not always realized following these rules (especially in spontaneous speech; Canepari, 1999). For example, an unemphatic trill [r] may occasionally and non-systematically reduce to a tap [ɾ], whereas a tap [ɾ] may be implemented as a trill [r] for emphasis (Canepari, 1999). Besides, other individual variants (e.g. uvular trill, alveolar approximant and fricative) may also be encountered (Celata et al., 2016; Romano, 2013). The /l/ in Standard Italian is an alveolar lateral approximant, which becomes a dentoalveolar when followed by /t, d, t͡s, d͡z, s, z/ and a palatalized postalveolar when followed by /t͡ʃ, d͡ʒ, ʃ/ (Canepari, 1999).

There is some debate on the nature of Mandarin Chinese /r/. In Lee and Zee (2003) and Lin (2007), it is described as a voiced postalveolar approximant [ɹ]; Duanmu (2007) describes it as a voiced retroflex fricative [ʐ]; Cerini (2013) states that a voiced postalveolar approximant [ɹ] is the primary rendition of Mandarin Chinese /r/, with [ʐ] being an important variant. Here, following Hall and Hamann (2010, p. 1833), we treat Mandarin Chinese /r/ as an approximant because 'First, it has phonetically little friction, and second, if treated as a fricative it would be the only voiced obstruent in Mandarin'. As for Mandarin Chinese /l/, it is usually described as a dental (Duanmu, 2007) or dentoalveolar lateral (Lee & Zee, 2003; Lin, 2007).

4.1.3 Theoretical framework

Among current L2 speech acquisition theories, only the Speech Learning Model (SLM; Flege, 1995) and its revised version (SLM-r; Flege & Bohn, 2021) explicitly deal with both L2 perception and production. Both models posit that the perceived acoustic dissimilarity between an L2 sound and the closest L1 sound(s) plays a vital role in L2 speech learning. Specifically, the more an L2 sound differs auditorily from the closest L1 sound(s), the easier it will be for L2 learners to discern cross-language phonetic differences and to form a new phonetic category for the L2 sound. Besides, according to both models, L2 speech perception and production are closely related, with the perception either preceding the production (SLM) or coevolving with the production (SLM-r).

In this study, we apply a different approach, namely PAM-AOH (the Perceptual Assimilation Model combined with the Articulatory Organ Hypothesis) proposed by Best et al. (2016) to Chinese learners' acquisition of Italian /r-l/ contrast. This is because both SLM and SLM-r assume that it is primarily the *acoustic* information that determines the acquisition of speech sounds. However, as noted in Yang & Chen (2019), the *acoustic* comparison may be more successful in predicting vowel than consonant (dis)similarity, since vowels are primarily characterized by common acoustic parameters (i.e. first and second formant frequencies), whereas consonants are not. Taking the lateral/trill contrast (/r-l/) as an example, while the acoustic signature of the lateral /l/ is an anti-formant near the frequency of $F3$ (Johnson, 2012), the trill /r/ is generally characterized by the alternation of closed and open phases on the spectrogram caused by tongue vibrations (Ladefoged & Maddieson, 1996). As such, the acoustic comparison between /r/ and /l/ can be rather complex because of the lack of common acoustic parameters. PAM-AOH, on the other hand, assumes that it is the *articulatory* information (i.e. articulatory organ, constriction location and constriction degree) that plays a fundamental role in speech learning. Since the articulatory information is less variable than acoustic events and can be readily compared across consonants (Yang & Chen, 2019), the predictions regarding Chinese learners' acquisition of Italian /r-l/ contrast made by PAM- AOH should be rather testable.

Besides, though PAM-AOH was initially developed to address infants' L1 speech perception and production development, it is 'likely to have broader relevance to advancing understanding of speech perception and production in late L2 learners' (Best et al., 2016, p. 252). In fact, its viability in L2 speech acquisition has been verified in Chinese learners' perceptual learning of Russian initial /r-l/ contrast (Yang & Chen, 2019). To contribute more necessary empirical evidence, we aim to find out whether this approach can also be extended to Chinese learners' acquisition of Italian /r-l/ contrast.

PAM-AOH divides consonant contrasts into three types, namely *between-organ* distinctions (i.e. two consonants are realized with different articulatory organs), *within-organ* distinctions (i.e. two consonants are realized with the same articulatory organ), and *privative organ* distinctions (i.e. a specified gesture of a certain articulatory organ is present in one consonant but absent in the other). The between-organ distinctions are easily discriminated throughout linguistic development, whether they are native or not. On the contrary, the discrimination between within-organ and privative organ distinctions is initially difficult for infants, but can improve with linguistic experience if the distinctions are native, and fails to improve if the distinctions are nonnative. In our case, the /r-l/ contrast in Italian is a within-organ distinction as both phonemes are produced by raising the tongue tip toward the alveolar ridge. On the other hand, Mandarin Chinese /r-l/ is a between-organ contrast as /l/ is produced by raising the tongue tip toward the alveolar ridge but /r/ mainly by raising the tongue dorsum toward the middle portion of the palate (Chen & Mok, 2021). So, the /r-l/ contrast in Italian is different from that in Mandarin Chinese and therefore nonnative for Chinese learners. Since, according to PAM-AOH, the perception of nonnative within-organ contrast declines with linguistic experience, it is possible that late Chinese learners of Italian may have difficulty differentiating perceptually between Italian /r-l/ contrast.

Moreover, since PAM-AOH was not originally developed to account for L2 acquisition, to extend it to L2 learning context, another factor, namely the influence of learners' L1, needs to be taken into consideration (Yang & Chen, 2019). In our case, the /l/ in Mandarin Chinese and Italian is highly similar (in syllable onset position) as it is commonly realized as a (dento)alveolar lateral approximant. By contrast, Mandarin Chinese /r/ is very different from Italian /r/ as the former is a postalveolar approximant realized by raising the tongue dorsum toward the middle portion of the palate, while the latter an alveolar trill realized by raising the tongue tip toward the alveolar ridge. Therefore, Chinese learners should be more familiar with Italian /l/ than with /r/. For this reason, we expect Chinese learners to perceive Italian /l/ better than /r/.

As for production, the core premise of PAM-AOH is that infants' speech production closely correlates with their perception. Specifically, it is young children's perceived articulatory information that guides their production. Since we assume that Chinese learners perceive Italian /l/ better than /r/, we expect them to correspondingly produce Italian /l/ better than /r/.

With regard to learning experience, PAM-AOH posits that the perceptual learning of nonnative within-organ distinctions can be gradually achieved with an increased learning experience for consonant contrasts of 'Category Goodness (CG) assimilation' (i.e. two nonnative consonants are assimilated to the same native consonant, yet with different fit degrees), but not for those of 'Single Category (SC) assimilation' (i.e. two nonnative consonants are assimilated to the same native consonant, each with a similarly high goodness-of-fit; PAM-L2; Best & Tyler, 2007). In our case, the /l/ in Italian and Mandarin Chinese is highly similar. By contrast, the similarity between Italian /r/ and Mandarin Chinese /l/ is relatively low; this is because though both phonemes are realized by raising the tongue tip toward the (dento)alveolar ridge, Italian /r/ features tongue tip vibration while Mandarin Chinese /l/ does not. Due to the difference in similarity extent, Chinese learners may assimilate Italian /r/ and /l/ to Mandarin Chinese /l/ with different degrees of fit, presumably forming a CG assimilation case. Thus, we speculate that Chinese learners' perception of Italian /r-l/ contrast will improve due to their increased learning experience. Consequently, guided by their enhanced perception, Chinese learners' production will improve as well.

4.1.4 Research questions and hypotheses

The research questions (RQ) addressed in this paper and the hypotheses (H) based on the above speculations are summarized as follows.

RQ1: How do Chinese learners perceive Italian /r-l/ contrast? Will their perception improve with their increased learning experience?

Hypotheses: Chinese learners will have difficulty differentiating between Italian /r-l/ contrast (H1), and will perceive Italian /l/ better than /r/ (H2). Besides, their perception will improve with their increased learning experience (H3).

RQ2: How do Chinese learners produce Italian /r-l/ contrast? Will their production improve with their increased learning experience?

Hypotheses: Guided by their perception, Chinese learners will produce Italian /l/ better than /r/ (H4). Besides, their production will improve as a result of their enhanced perception (H5).

RQ3: What is the relationship between Chinese learners' perception and production of Italian /r-l/ contrast?

Hypothesis: The two speech modalities will be closely related in Chinese learners' Italian /r-l/ contrast acquisition (H6).

To address the above research questions and verify the hypotheses, we carried out a perception and a production experiment.

4.2 Methods

4.2.1 Participants

The Chinese participants in the perception and production experiments were 10 first-year (1MC group: Female = 9, Male = 1, Mean age = 19.1), 10 second-year (2MC group: Female = 10, Male = 0, Mean age = 19.9) and 10 third-year (3MC group: 10 Female = 8, Male = 2, Mean age = 20.9) Chinese undergraduate students majoring in Italian at Dalian University of Foreign Languages in China. The experiments were conducted at the end of the academic year, which means that at the time of the experiments, the 1MC, 2MC and 3MC groups had studied Italian for about one year, two years and three years, respectively. The reason for using the learning length as the grouping criterion was to investigate the effect of learning experience on Chinese participants' acquisition process. Note also that none of the Chinese participants had had a study abroad experience in Italy at the time of the experiments. The Chinese participants' gender distribution reflects the imbalance of the students' enrollment in the degree course. Besides, 10 native Italian-speaking undergraduate students (NIT group: Female = 7, Male = 3, Mean age = 21.8) from the Veneto region in the North-East of Italy served as controls. No participant reported any hearing or speech impairment at the time of the experiments.

In addition, three native Italian-speaking postgraduate students were recruited as raters of the above participants' productions. The three raters comprised one female (age = 23) from the Veneto region in the North-East of Italy, one male (age = 24) from the Abruzzo region in central Italy and one male (age = 24) from the Puglia region in the south of Italy. The reason for recruiting raters from the three different parts of Italy was to minimize the dialectal influence on their ratings.

4.2.2 Stimuli

The word stimuli used in the perception and production experiments were the same. This was done to maintain the consistency between the two experiments and thus facilitate the testing of our hypothesis regarding the relationship between the two speech modalities. Specifically, the target word stimuli were six Italian minimal pairs contrasting in /r-l/ in three different phonetic contexts (i.e. 2 in word-initial prevocalic position (#cV), 2 in word-internal intervocalic position (VcV), and 2 in word-internal preconsonantal position (VcC); see Appendix III). Besides, ten Italian disyllabic minimal pairs contrasting in stop voicing (i.e. voiced vs. voiceless) served as distractors. The use of a higher number of distractors than target stimuli was to prevent the participants from grasping the study purpose. The carrier sentence was *Leggo* ___ *bene* 'I read ___ well'.

4.2.3 Material preparation

For the perception experiment, all word stimuli were first ordered randomly in isolation, then embedded in the carrier sentences and arranged in random order. The reason for including these two types of stimuli (i.e. in isolation vs. in the carrier sentences) was to bring the experiment environment close to real life, since the words heard (and spoken) in daily communication are not always in isolation but more often in sentences. Afterwards, the stimuli in isolation and in the carrier sentences were carefully read by a female native Italian speaker (age = 24) from the Veneto region. The recording was collected in the Language and Communication Lab at the University of Padova in Italy, using a Roland R09 voice recorder with a sampling rate of 44.1 kHz and 16-bit resolution. In total, 64 trials (16 minimal pairs × 2 words × 2 contexts = 64) were produced; their clarity and correctness were verified auditorily by a native Italian-speaking phonetician (i.e. the second author). Finally, the trials were extracted from the recording and saved separately as isolated words or sentences. The intensity of the separate audio files was normalized to 70 dB.

For the production experiment, the word stimuli were first repeated twice randomly in isolation, then inserted in the carrier sentences and repeated twice in random order. The stimuli in isolation and in the carrier sentences were printed on a paper sheet for the participants to read. After every 10 words/sentences, there was a cartoon picture to remind the participants to have a short break.

4.2.4 Procedure

The perception experiment was conducted before the production experiment. The Chinese learners participated in both experiments in a multimedia speech lab at Dalian University of Foreign Languages in China, and the native Italian participants in the Language and Communication Lab at the University of Padova in Italy.

a. Perception experiment

A perceptual identification test was developed with the above-mentioned audio files, using the Alchemer platform (www.alchemer.com/). In the test, the 32 isolated words, followed by the 32 sentences, were randomly presented to the participants. Each word/sentence was accompanied by two response options (i.e. a minimal pair). The participants, wearing headphones and sitting in front of a computer, were instructed to click on the option given on the screen that corresponded to the word heard or that appeared in the sentence heard. Each word/sentence was repeated only once. The participants were allowed to listen to each word/sentence for a maximum of three times when it was presented. The test lasted about 10 minutes.

b. Production experiment

First, The participants were given a paper sheet and instructed to read the words and sentences on it in a natural way. They were also asked to make a short break when the cartoon pictures told them to. The recordings of the Chinese learners were administered using a Zoom H4n Pro voice recorder with a sampling rate of 44.1 kHz and 16-bit resolution, and those of the native Italian speakers using a Roland R09 voice recorder with a sampling rate of 44.1 kHz and 16-bit resolution. In total, 1920 target tokens (6 target minimal pairs × 2 words × 2 contexts × 2 repetitions × 40 participants = 1920) were elicited. The target tokens were extracted from the above recordings and saved separately as isolated words or sentences. The intensity of the separate tokens was equalized to 70 dB.

Subsequently, three rating tasks were developed with the above tokens as stimuli using ExperimentMFC 7 in Praat (Boersma & Weenink, 2020). Table 15 shows the stimulus composition of the three rating tasks. As can be seen, each task had 768 stimuli (6 minimal pairs × 2 words × 2 contexts × 2 repetitions × [12+4] participants = 768); 75% (12/[12+4] participants = 75%) of the stimuli were different across tasks, and the remaining 25% were identical. Each task consisted of four parts. In the first part, 192 isolated words (3 minimal pairs × 2 words × 2 repetitions × 16 participants = 192) were randomly presented to the raters. Meanwhile, 6 words (i.e. the 3 minimal pairs) were shown on the screen as options. Besides, a *Nessuna* 'none' option and an 'oops' button were also given. In the second part, the remaining 192 isolated words were presented. The third and fourth parts were structured similarly to the first two, except that the stimuli were the 384 sentences.

Table 15. Stimulus composition of the three rating tasks.

Task	Different stimuli (i.e. tokens produced by participants)	Identical stimuli (i.e. tokens produced by participants)
1	1MC-1, 1MC-2, 1MC-3, 2MC-1, 2MC-2, 2MC-3, 3MC-1, 3MC-2, 3MC-3, NIT-1, NIT-2, NIT-3.	1MC-10, 2MC-10, 3MC-10, NIT-10.
2	1MC-4, 1MC-5, 1MC-6, 2MC-4, 2MC-5, 2MC-6,	

	3MC-4, 3MC-5, 3MC-6, NIT-4, NIT-5, NIT-6.
3	1MC-7, 1MC-8, 1MC-9, 2MC-7, 2MC-8, 2MC-9, 3MC-7, 3MC-8, 3MC-9, NIT-7, NIT-8, NIT-9.

Finally, the three raters were randomly assigned to one of the three tasks. In the first two parts, the raters, seated in front of a computer with headphones on, were asked to click on the option corresponding to the word heard. If they found that the word heard did not correspond to any of the 6 options, they could click on the option of *Nessuna* 'none'. If they accidentally clicked on the undesired option, they could go back to the previously heard stimulus by clicking on the 'oops' button. In the last two parts, the raters performed similarly as in the first two parts, except that they had to determine which of the 6 words given on the screen appeared in the sentences heard. Each word/sentence was repeated only once. The raters could listen to each word/sentence up to three times in case sometimes they might need more than one repetition to identify what they heard. After every 32 words/sentences there was a break. The task was self-paced and generally lasted about 30 minutes. At the final result extraction stage, the *Nessuna* 'none' responses were automatically extracted as '0'; the responses containing target /r/ (e.g. *rana* 'frog') were extracted as 'r', and those containing target /l/ (e.g. *lana* 'wool') as 'l'.

4.2.5 Data preparation and analyses

a. Perception experiment

A generalized linear mixed-effects model (GLMM) with a binomial link function was applied to the participants' responses (i.e. correct vs. incorrect) using the lme4 package 1.1.26 (Bates et al., 2015) in R 3.6.3 (R Core Team, 2020). The GLMM had Group (four levels: 1MC, 2MC, 3MC, NIT), Consonant (two levels: /r/ and /l/), and the interaction between them as fixed factors, and Subject and Item as random intercepts. The coding scheme used was treatment coding. The main effects of the fixed factors were assessed with the Type II Wald chi-squared tests using the car package 3.0.10 (Fox & Weisberg, 2019). Post-hoc pairwise comparisons with FDR (false discovery rate) correction were implemented using the emmeans package 1.5.3 (Lenth, 2020).

b. *Production experiment*

First of all, on the basis of the three raters' responses to the 192 stimuli (768 stimuli × 75% = 192) shared by the three rating tasks, the interrater reliability was calculated using the irr package 0.84.1 (Gamer et al., 2019) in R 3.6.3 (R Core Team, 2020).

In the second place, we determined the final responses to the shared stimuli as all three raters assessed them, and we thereby had three different sets of responses. The determination was made following the 'majority rule' when there was divergence in the raters' responses. For example, if a stimulus was identified as 'r' by two raters and as 'l' by one rater, its final response would be 'r'. Note that none of the stimuli was identified as '0', 'r', and 'l' respectively by the three raters. In total, only 29 (out of 192) stimuli required the application of the 'majority rule'; the responses to the other 163 stimuli (192–29 = 163) were identical across the three raters.

Afterwards, an auditory analysis was performed using praat on the tokens that were classified as '0' (*Nessuna* 'none'). Specifically, we identified these tokens auditorily and transcribed them into corresponding (non-)words. Based on the transcription, we divided the tokens into three broad categories depending on the reasons for which they were not recognized as any of the 12 given words by the three raters. In the first category, all the tokens were produced as real words, but different from the ones that were given. For example, *regno* 'kingdom' was misread as *ragno* 'spider'; *farsi* 'make oneself' as *frasi* 'sentences'. These tokens were labeled 'other real words'. In the second category, all the tokens were mispronounced as non-words and therefore did not correspond to any given word. Moreover, the mispronounced phonemes were other than the target /r/ and /l/. For example, *farsi* ‘make onself’ was mispronounced as *forsi**, *Parma* ‘city name’ as *Barma**. This category was labeled ‘non-words with no target errors’. The third category was labeled ‘non-words with target errors’. It included all the tokens that were also mispronounced as non-words, yet, the mispronunciation occurred on the target /r/ or /l/. Some examples of this category can be seen in the following Results section.

Finally, the ‘0’ responses were discarded from the statistical analyses and the remaining responses were recoded. Specifically, if the raters’ responses matched the target consonants in the stimuli, the responses would be recoded as ‘correct’ (e.g. the ‘r’ response to stimulus *rana* was recoded as ‘correct’); otherwise, they would be recoded as ‘incorrect’ (e.g. the ‘r’ response to stimulus *lana* was recoded as ‘incorrect’). A GLMM similar to that applied to the participants’ responses in the perception experiment was applied to the raters’ responses (i.e. ‘correct’ vs. ‘incorrect’) in the production experiment. The only difference is the inclusion of a random slope for Consonant by Subject in the random effects structure.

4.3 Results

4.3.1 Perception experiment

We obtained a total of 960 responses for the target stimuli (6 target minimal pairs × 2 words × 2 contexts × 40 participants = 960). The participants’ perceptual accuracy rates calculated based on these responses are illustrated in Figure 9. As the figure shows, the NIT group’s perceptual identification accuracy for both Italian /r/ and /l/ reached the ceiling. By contrast, the accuracy of the three Chinese learner groups was relatively lower, yet far above chance level. Besides, all Chinese learner groups had similar perceptual accuracy for Italian /r/ and /l/, and their perception did not appear to improve with their increased learning experience.

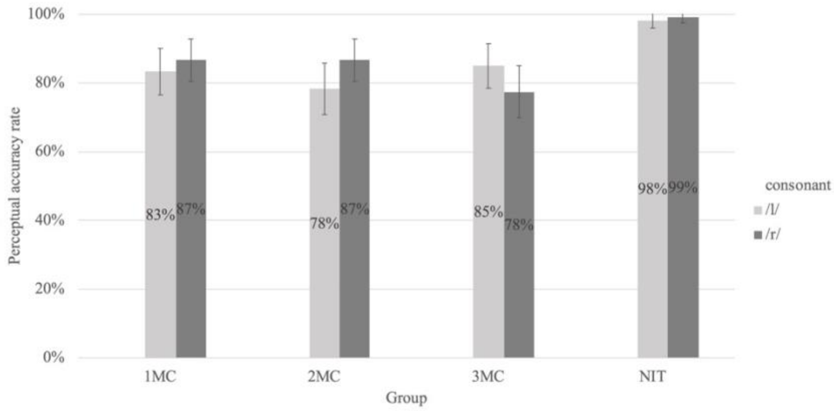


Figure 9. 1MC, 2MC, 3MC and NIT groups' perceptual accuracy rates for Italian /r-l/ contrast.

The results of the GLMM applied to the participants' responses in the perceptual identification test are shown in Table 16. As can be seen, the GLMM yielded a significant main effect on Group, indicating that the participants' perceptual accuracy differed significantly by group. Post-hoc pairwise comparisons revealed that there were no significant differences across the three learner groups (1MC vs. 2MC: $\beta = –0.20$, $SE = 0.65$, $z = –0.31$, $p = 0.85$; 1MC vs. 3MC: $\beta = –0.32$, $SE = 0.65$, $z = –0.50$, $p = 0.85$; 2MC vs. 3MC: $\beta = –0.12$, $SE = 0.64$, $z = –0.19$, $p = 0.85$), which were all significantly less accurate than the NIT group (1MC vs. NIT: $\beta = 2.95$, $SE = 0.91$, $z = 3.26$, $p = 0.0022$; 2MC vs. NIT: $\beta = 3.15$, $SE = 0.91$, $z = 3.47$, $p = 0.0015$; 3MC vs. NIT: $\beta = 3.27$, $SE = 0.90$, $z = 3.63$, $p = 0.0015$). Besides, the nonsignificant interaction between Group and Consonant indicates that the ways in which the identification accuracy differed between groups did not vary by consonant. Moreover, the nonsignificant main effect of Consonant indicates that the participants' identification accuracy did not vary by this factor.

Table 16. Results of the GLMM of the participants' perception of Italian /r-l/ contrast (p values < 0.05 are in bold).

	Fixed effects			Random effects	
				By Subject	By item
	χ^2	*df*	*p*	*SD*	*SD*
Intercept	-	-	-	1.55	0.60
Group	15.04	3	**0.002**	-	-
Consonant	0.01	1	0.94	-	-
Group × Consonant	7.06	3	0.07	-	-

4.3.2 Production experiment

The interrater reliability was high (Fleiss k = 0.802), which means that the three raters, though from different parts of Italy, substantially agreed with each other in perceptually identifying Italian /r-l/ contrast.

Out of the 1920 target tokens, only 66 were rated as *Nessuna* 'none' (i.e. did not correspond to any of the given response options). According to the results of the auditory analysis, these tokens were distributed as follows: 10 belonged to the first category ('other real words'), 26 to the second category ('non-words with no target errors'), and 30 to the third category ('non-words with target errors'). We were particularly interested in the third category, and a closer inspection of the tokens in it revealed that the mispronunciation errors of the target /r/ and /l/ can be further divided into four types, namely nasalization (i.e. production of /r/ or /l/ as nasal /n/; e.g. *lana* 'wool' as *nana**), gemination (i.e. production of singleton /l/ as geminate /l:/; e.g. *male* 'bad' as *malle**), epenthesis (i.e. insertion of redundant phonemes around the target /r/ or /l/; e.g. *cara* 'dear' as *catra**), and deletion (i.e. complete omission of the target /r/ or /l/; e.g. *palma* 'palm' as *pama**). Among the 30 tokens in this category, we found 12 cases of nasalization, 9 of gemination, 7 of epenthesis, and 2 of deletion.

The vast majority ([1920-66]/1920 = 96.6%) of Italian /r-l/ contrast was realized either as /r/ or as /l/ by the participants. Based on the three native Italian listeners' perceptual ratings, the four groups' production accuracy rates for Italian /r-l/ contrast were calculated and are shown in Figure 10. As the figure shows, the vast majority of the /r-l/ contrast produced by the native Italian controls was realized as intended. Regarding the Chinese learners, all three groups had higher production accuracy for Italian /l/ than for /r/. Moreover, while the Chinese learners' production accuracy of Italian /r/ improved with their increased learning experience, the accuracy of /l/ decreased.

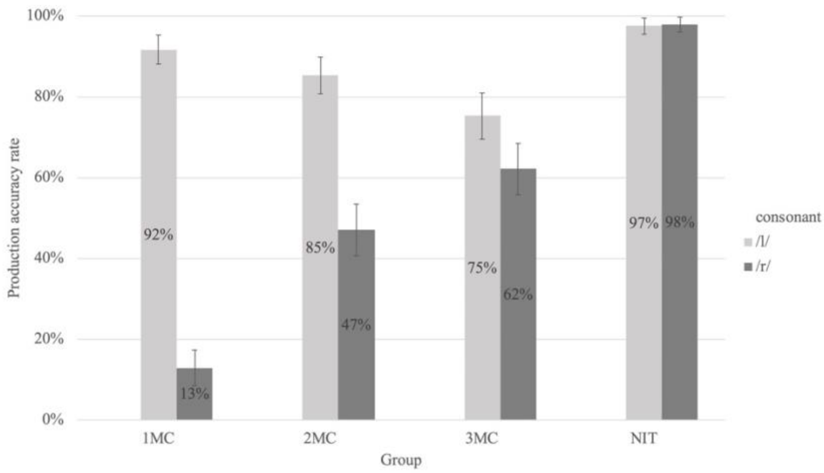


Figure 10. 1MC, 2MC, 3MC and NIT groups' production accuracy rates for Italian /r-l/ contrast.

Table 17 shows the results of the GLMM applied to the raters' responses in the rating task. As can be seen, the GLMM yielded significant main effects on all fixed factors, namely Group, Consonant, and the interaction between them. This indicates that the participants' production accuracy differed significantly not only by group, but also by consonant. Besides, the production accuracy differences between groups varied by consonant.

Table 17. Results of the GLMM of the participants' production of Italian /r-l/ contrast (p values < 0.05 are in bold).

	Fixed effects			Random effects	
				By Subject	By item
	χ^2	*df*	*p*	*SD*	*SD*
Intercept	-	-	-	0.94	0.39
Group	41.54	3	**<.001**	-	-
Consonant	10.37	1	**0.001**	9.91	-
Group × Consonant	18.63	3	**<.001**	-	-

The results of the post-hoc pairwise comparisons are listed in Table 18. First, as for Italian /r/, all three Chinese learner groups were significantly different from the NIT group, indicating that the Chinese learners never reached the native level in producing Italian /r/. Besides, both the 2MC and 3MC groups were significantly more accurate than the 1MC group, indicating that the Chinese learners' increased learning experience positively affected their production of Italian /r/. Second, regarding Italian /l/, the 1MC group was not different from the NIT group, which means that the Chinese learners with one year of learning experience had no difficulty producing Italian /l/. However, the 2MC and 3MC groups differed significantly from the NIT group. Also, there was a significant difference between the 1MC and 3MC groups. These results indicate that the Chinese learners' increased learning experience had a negative impact on their Italian /l/ production accuracy.

Table 18. Results of the post-hoc pairwise comparisons of the 1MC, 2MC, 3MC and NIT groups' production of Italian /r-l/ contrast (p values < 0.05 are in bold).

	Estimate	*SE*	*z* ratio	*p* value
Consonant /r/				
1MC vs. 2MC	2.69	1.23	2.19	**0.0347**
1MC vs. 3MC	3.81	1.24	3.07	**0.0032**
2MC vs. 3MC	1.12	1.20	0.93	0.35
1MC vs. NIT	8.34	1.45	5.74	**<0.001**
2MC vs. NIT	5.65	1.38	4.09	**0.0001**
3MC vs. NIT	4.53	1.36	3.32	**0.0018**
Consonant /l/				
1MC vs. 2MC	–0.78	0.56	–1.39	0.20
1MC vs. 3MC	–1.44	0.55	–2.60	**0.0185**
2MC vs. 3MC	–0.66	0.52	–1.27	0.20
1MC vs. NIT	1.31	0.68	1.93	0.08
2MC vs. NIT	2.09	0.66	3.17	**0.0046**
3MC vs. NIT	2.75	0.65	4.21	**0.0002**

4.4 Discussion and conclusions

This study set out to investigate how L1-Chinese learners with different learning experiences perceive and produce the /r-l/ contrast in Italian.

Regarding the first research question, addressing how Chinese learners perceive Italian /r-l/ contrast and whether their perception improves with their increased learning experience, our results show that, first, the Chinese participants had significantly lower perceptual identification accuracy than the native Italian listeners, suggesting that Chinese learners have difficulty differentiating between Italian /r-l/ contrast. This is consistent with our H1, indicating that the perceptual discrimination between nonnative within-organ consonant contrast represents a challenge for late L2 learners, as assumed by PAM-AOH (Best et al., 2016).

Second, our results do not support our H2. That is, the Chinese learners did not perceive Italian /l/ better than /r/, though they were much more familiar with the former than with the latter as /l/ is highly similar in Italian and Mandarin Chinese. This may be due to the influence of Mandarin Chinese phonotactic rules. Specifically, in Mandarin Chinese /l/ can only occur in syllable onset position, while this perception experiment included Italian stimuli having /l/ in syllable coda position (i.e. *palma* 'palm', *falsi* 'fake'). We assume that when two phonetically similar sounds (one of which is familiar, the other is not; i.e. /l/ and /r/) are found in a position (i.e. syllable coda) where the familiar sound (i.e. /l/) cannot occur according to the L1 phonotactic rules, Chinese learners may tend to categorize both sounds as unfamiliar (i.e. /r/). In this way, the Chinese learners' better perception of Italian /l/ was mitigated by the influence of their L1 phonotactics, resulting in a similar perceptual accuracy for Italian /l/ and /r/. In short, we speculate that Chinese learners' L1 phonotactic constraints interfere in their L2 sound processing. However, it should be noted that at the present stage this is a pure speculation. To verify it, a future perception experiment with a larger number of Italian word stimuli with /r-l/ contrasting in various syllable positions is needed.

Third, contrary to our H3, the three Chinese learner groups with different learning experiences had similar perceptual identification accuracy for Italian /r-l/ contrast, suggesting that Chinese learners' perception fails to improve with their increased learning experience. There are two possible explanations for this. The first one is that the Chinese learners' perception of Italian /r-l/ contrast did not form a 'Category Goodness (CG) assimilation' case as we expected, but formed a 'Single Category (SC) assimilation' one where the learners assimilated both Italian /l/ and /r/ to Mandarin Chinese /l/ with similar high fit degrees. Since according to PAM-AOH the perceptual learning of nonnative within-organ consonant distinctions of 'SC assimilation' case cannot be achieved with an increased learning experience, the Chinese learners' increase in their Italian learning length was not accompanied by an enhancement in their perception of Italian /r-l/ contrast.

Alternatively, it is possible that the Chinese learners did form a 'CG assimilation' case in the perception of Italian /r-l/ contrast. This is because the Chinese learners' perceptual identification accuracy for both Italian /r/ and /l/ was always well above chance level (78%~87%), and only the 'CG assimilation' but not the 'SC assimilation' case can lead to discrimination performance ranging from fair to good (PAM-L2; Best & Tyler, 2007). Yet, why the Chinese learners' perception of Italian /r-l/ contrast did not improve with their increased learning experience as predicated by PAM-AOH? This may be because at the time of the experiment, even the Chinese learners with the least learning experience had studied Italian for one year. As Best and Tyler (2007) indicated, the cut-off of L2 perceptual learning should be six to twelve months. After this initial period, additional experience seems no longer to enhance L2 perception for most late learners. For this reason, late L2 learners with 1.5 to 5 years of learning experience do not always outperform those with 6-12 months of experience (Best & Tyler, 2007). As such, in the present perception experiment, the Chinese learners with only one year of learning experience were as accurate as those with two/three years of experience in identifying perceptually Italian /r-l/ contrast. Similarly, the null effect of learning experience on the /r-l/ perception by Chinese learners with more than one year of learning experience has also been reported in their L2 European Portuguese (Zhou, 2021) and Russian (Yang & Chen, 2019) learning.

Regrettably, at the present stage we cannot ascertain which of the above two explanations holds for the results of the present perception experiment. To address it, a future perceptual assimilation task (PAT) is needed.

As for the second research question, addressing how Chinese learners produce Italian /r-l/ contrast and whether their productions will improve with their increased learning experience, our results show that, in line with the H4, the Chinese learners produced Italian /l/ better than /r/. This is not surprising as the lateral approximant /l/ is available in Mandarin Chinese phonology, and a positive transfer from L1 to L2 could result in Chinese learners' successful realization of Italian /l/. By contrast, the trill /r/ is absent from Mandarin Chinese phonological system, and is therefore a completely novel sound to Chinese learners. Furthermore, the realization of the trill /r/ is rather challenging as 'The aperture size and airflow must fall within critical limits for trilling to occur, and quite small deviations mean that it will fail' (Ladefoged & Maddieson, 1996, p. 217). Due to this, the Chinese learners experienced difficulty in successfully producing Italian /r/. Besides, almost half or more of the Italian /r/ produced by the Chinese learners was perceived as /l/ by the native Italian raters, indicating that when failing to produce Italian trill /r/, Chinese learners tend to use the lateral /l/ as a substitute. This repair strategy has also been reported in Chinese learners' Spanish (Patience, 2018), European Portuguese (Zhou et al., 2021), and Persian (Falahati, 2015) learning.

Regarding the acquisition development, our results show that the more experienced the Chinese learners became, the more times they could successfully produce Italian /r/. This is consistent with our H5, suggesting that Chinese learners' increased learning experience does favor their Italian /r/ acquisition. The positive effect of learning experience was also found in Chinese learners' Spanish /r/ production (Patience, 2018). Besides, it should be noted that the Chinese learners' proper /l/ production occurrences decreased with their increased learning length. In other words, the more experienced the Chinese learners were, the more often they mispronounced Italian /l/ as /r/. We attribute this fact to the Chinese learners' hypercorrection. Specifically, the Chinese learners might be aware of their erroneous replacement of Italian /r/ by /l/ in production, and the approach they used to correct this error was to overproduce /r/ even when /l/ rather than /r/ was needed. This hypercorrection was not obvious in the case of the less experienced Chinese learners as they had difficulty successfully producing Italian /r/. However, as the learning experience increased and the Chinese learners became more capable of producing Italian /r/, this hypercorrection became more evident.

Regarding the third research question, addressing the relationship between Chinese learners' perception and production of Italian /r-l/ contrast, our results did not show consistency between how the Chinese learners' perceived and produced Italian /r-l/ contrast. Specifically, the Chinese learners did not identify perceptually Italian /l/ better than /r/, but their production of Italian /l/ was better than that of /r/. Moreover, while the Chinese learners' perceptual identification accuracy of Italian /r-l/ contrast stayed constant, their production varied with their increased learning experience. These results, taken together, suggest that contrary to our H6, Chinese learners' perception may not be closely related to their production.

We argue that the lack of a strong correlation between the two speech modalities in Chinese learners' Italian /r-l/ contrast acquisition is due to the dissociations between the perception and production in L2 speech learning. According to Ramus et al. (2010), there is a distinction between the speakers' perception and production grammars. The speakers' perception 'starts from an auditory phonetic representation and aims at arriving at an underlying lexical-phonological representation', whereas their production 'starts from an underlying lexical-phonological representation and aims at arriving at an articulatory phonetic representation' (Boersma & Hamann, 2009, p. 1; Ramus et al, 2010). These two grammars develop parallelly in L1 children (as assumed by PAM-AOH) and ultimately become indistinguishable in monolingual adults, but can diverge in late L2 learners (Ramus et al., 2010). Possibly due to this, accurate L2 perception cannot always lead to accurate L2 production, and the two speech modalities in late L2 learners may follow two independent development pathways such that while the learners' production can be modified by increased exposure to L2 input or improved grammatical proficiency, this may not be the case for their perception (de Leeuw et al., 2021). Furthermore, the dissociations between the perception and production in L2 speech learning have been supported by brain image studies (Golestani et al., 2007; Golestani & Pallier, 2007), and have also been observed in other empirical studies (e.g. Hattori & Iverson, 2010; Kartushina & Frauenfelder, 2014; Kassaian, 2011; Kluge et al., 2007; Neufeld, 1979; Peperkamp & Bouchon, 2011; Sheldon & Strange, 1982). Therefore, we argue that the core premise of PAM-AOH, namely that L1 infants' production is primarily guided by their perceived articulatory information, may not account for the relationship between the two modalities in late L2 speech learning. However, considering that the present study focused solely on Italian /r-l/ contrast acquisition and included a relatively restricted number of participants and stimuli, to consolidate this argument, more empirical evidence from the cross-linguistic acquisition of other consonant contrasts is needed.

The core premise of PAM-AOH, nevertheless, triggers an interesting question regarding pronunciation training. Specifically, according to PAM-AOH, L1 infants' perception of speech articulatory information can be multimodal. That is, L1 infants can detect the articulatory gesture information non only through hearing the speech (auditory information), but also via seeing the talking face (visible information). Since a growing body of research has noted that computer assisted pronunciation training (CAPT) which provides visible information about articulatory configurations, movements, and processes can favor L2 speech acquisition (see Bliss et al. [2018] for a review), it is interesting to know whether CAPT that can *visually* present the articulatory gestures of Italian /r-l/ contrast to Chinese learners will help them to better capture the articulatory information of these two phonemes and consequently improve their acquisition.

To conclude, similar to the /r-l/ contrast in other languages such as Spanish, European Portuguese, Russian and Persian, Italian /r-l/ contrast represents an acquisition challenge for L1-Chinese learners. Specifically, in perception, Chinese learners have difficulty differentiating between Italian /r-l/ contrast. However, they do not confound these two phonemes completely, but can distinguish them to a certain extent. In production, Chinese learners have more difficulty successfully realizing Italian /r/ than /l/, and they tend to replace the rhotic /r/ with the lateral /l/ when failing to produce the former. Also, while Chinese learners' proper production occurrences of Italian /r/ and /l/ increase and decrease, respectively, with the growth in their learning experience, their perceptual identification accuracy of Italian /r-l/ contrast remains unchanged. In other words, Chinese learners' perception and production of Italian /r-l/ contrast do not appear to develop in a parallel fashion, suggesting a possible dissociation between the two modalities in L2 speech acquisition.

Chapter 5

Conclusions and Final Remarks

This doctoral dissertation aimed to empirically investigate the understudied issue of how Mandarin Chinese-speaking learners acquire Italian consonant sounds. More specifically, it looks at Chinese learners' acquisition of Italian voiced vs. voiceless stop consonant contrast, singleton vs. geminate consonant contrast, and lateral vs. rhotic consonant contrast, which have been documented to be problematic for Chinese learners. The main findings are summarized as follows.

First, regarding stop consonants, Chinese learners have difficulty differentiating perceptually between Italian voiced and voiceless stops as they categorize both types of stops within a single category. In production, Chinese learners have difficulty producing properly Italian voiced (but not voiceless) stops, which are absent from the Mandarin Chinese phonology. When failing to produce Italian voiced stops, Chinese learners' repair strategy is to replace them with Mandarin Chinese voiceless unaspirated stops.

Second, as for consonant length contrast, Chinese learners can distinguish perceptually between Italian singleton and geminate consonants to a certain degree, but not in a native-like manner. Besides, though Chinese learners are more familiar with singleton than geminate consonants, they do not perceive Italian singletons more accurately than geminates. In production, Chinese learners succeed in making a distinction between shorter singleton and longer geminate consonants in Italian. However, their geminate to singleton consonant duration ratios are much smaller than native Italian speakers'. Also, the duration interplay between Italian consonants and preconsonantal vowels (i.e., longer vowels before singleton consonants and shorter vowels before geminate consonants) is entirely ignored by Chinese learners. Moreover, an increased learning experience does not appear to enhance Chinese learners' acquisition of Italian consonant length contrast.

Third, concerning Italian /r-l/ contrast, in perception Chinese learners have some difficulty differentiating these two phonemes; and despite the fact that Chinese learners are much more familiar with the lateral /l/ than with the rhotic /r/ due to the former being present in Mandarin Chinese phonology, they do not perceive Italian /l/ better than /r/. Also, Chinese learners' perceptual accuracy remains unchanged even when their learning experience increases. In production, Chinese learners have more difficulty realizing properly Italian /r/ than /l/, and show the tendency to replace /r/ with /l/. However, while Chinese learners' production of Italian /r/ improves with their increased learning experience, their accurate /l/ production occurrences decrease.

Overall, the above results show that the acquisition of Italian stop contrast, consonant length contrast, and /r-l/ contrast is indeed challenging for Chinese learners, as documented by the previous descriptive studies reviewed in §1.4. Importantly, as compared to those descriptive studies, this dissertation provides a much more detailed and complete picture of how Chinese learners acquire the three consonant contrasts in Italian. Nonetheless, the present work has certain limitations. First of all, since all three empirical studies were conducted during the Covid-19 pandemic, the number of participants recruited for each study was relatively restricted because of the restrictions on gatherings imposed by the pandemic. Future replication studies with a larger number of participants are needed to consolidate the current conclusions. In the second place, the perception and production experiments of each study were conducted in a controlled fashion. Future studies might want to include listening/reading tasks conducted in a less controlled and more naturalistic setting. Lastly, only cross-sectional studies are included in this dissertation. To better analyze the role played by learning experience in the developmental acquisition of Italian consonant sounds by Mandarin Chinese-speaking learners, longitudinal studies are needed.

This doctoral dissertation also leaves several avenues for future research. Specifically, it has been shown that substantial between-subject differences exist in L2 speech learning (e.g., Chandrasekaran et al., 2010; Darcy et al., 2015; Inceoglu, 2019; Sun & Zhang, 2020). Therefore, future research could pay more attention to Chinese learners' individual differences in their Italian consonant sound acquisition.

Besides, considering that computer-assisted pronunciation training (CAPT) can play a positive role in novel segmental feature acquisition (see Bliss et al. [2018] for a review), future research might want to investigate whether this approach can aid Chinese learners of Italian to better master the three challenging consonant contrasts in analysis here. Also, though this dissertation fills some research gaps, the acquisition of L2 Italian speech by Mandarin Chinese-speaking learners remains a much understudied subject. To obtain a complete picture of this acquisition process, future research should extend the focus not only to other consonants and vowels, but also to the suprasegmental domain (e.g., stress, prosody).

Bibliography

Adda-Decker, M., & Lamel, L. (2005). Do speech recognizers prefer female speakers? *Proc. INTERSPEECH 2005*, 2205–2208.

Altmann, H., Berger, I., & Braun, B. (2012). Asymmetries in the perception of non-native consonantal and vocalic length contrasts. *Second Language Research*, *28*(4), 387–413. https://doi.org/10.1177/0267658312456544

Aoyama, K. (2000). *A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity distinctions*. University of Hawai'i at Manoa.

Bassetti, B., Sokolović-Perović, M., Mairano, P., & Cerni, T. (2018). Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Language and Speech*, *61*(4), 577–597. https://doi.org/10.1177/0023830918780141

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. True voice languages. *Journal of Linguistics*, *49*, 259–284.

Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131–151. https://doi.org/10.1017/S0025100305002148

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). John Benjamins.

Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating What Infants Attune to in Native Speech. *Ecological Psychology*, *28*(4), 216–261. https://doi.org/10.1080/10407413.2016.1230372

Bliss, H., Abel, J., & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, *4*(1), 129–153. https://doi.org/10.1075/jslp.00006.bli

Boersma, P., & Hamann, S. (2009). Introduction: Models of phonology in perception. In P. Boersma & S. Hamman (Eds.), *Phonology in perception* (Vol. 15, pp. 1–24). Walter de Gruyter.

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (6.1.16) [Computer software]. http://www.praat.org/

Bortolini, U., Zmarich, C., Fior, R., & Bonifacio, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology*, *31*(2–3), 191–206.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *The Journal of the Acoustical Society of America*, *63*(5), 1556–1558.

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136–193. https://doi.org/10.1191/026765898669508401

Calabrò, L., & Mairano, P. (2017). Insegnare la pronuncia a studenti sinofoni di italiano L2 con il Minimal Pair Finder. *Italiano Lingua Due*, *9*(1), 83–97.

Canepari, L. (1999). *Il MaPi: Manuale di pronuncia italiana*. Zanichelli.

Canepari, L., & Giovannelli, B. (2008). La buona pronuncia italiana del terzo millennio. In *Manualetto d'italiano neutro con sonori, esercizi e test (4a edizione)*. Aracne.

Cao, J. (2018). *Italiano L2 per apprendenti sinofoni: Problemi e proposte [L2 Italian for synophonic learners: Problems and proposals]* [Unpublished master's thesis]. Università degli Studi di Padova.

Cao, Q. (2018). *Perceção das Consoantes Líquidas por Aprendentes Chineses do Português Língua Estrangeira [Perception of Liquid Consonants by Chinese Learners of Portuguese as a Foreign Language]* [Unpublished master's thesis]. University of Aveiro.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (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*(2), 421–428.

Cavanaugh, J. R. (2005). Accent matters: Material consequences of sounding local in northern Italy. *Language & Communication*, *25*(2), 127–148. https://doi.org/10.1016/j.langcom.2005.02.002

Celata, C., Vietti, A., & Spreafico, L. (2016). An articulatory account of rhotic variation in Tuscan Italian: Synchronized UTI and EPG data. In M. Gibson & J. Gil (Eds.), *Romance Phonetics and Phonology* (pp. 91–117). Oxford University Press.

Cerini, M. (2013). *The pronunciation of Mandarin Chinese according to the canIPA natural phonetics & tonetics method* [Unpublished doctoral dissertation]. Ca' Foscari University of Venice.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465. https://doi.org/10.1121/1.3445785

Chao, K., & Chen, L. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics & Chinese Language Processing*, *13*, 215–232.

Chen, S., & Mok, P. P. K. (2021). Articulatory and Acoustic Features of Mandarin /ɹ/: A Preliminary Study. *12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 1–5. https://doi.org/10.1109/ISCSLP49672.2021.9362070

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

Chih, M. T.-C. (2013). E/LE en Taiwán: Problemas de apreciación fonética en estudiantes universitarios de grado [E/LE in Taiwan: Phonetic Assessment Problems in Undergraduate University Students]. *SinoELE*, *9*, 17–32.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229. https://doi.org/10.1006/jpho.1999.0094

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), 1–25. https://doi.org/10.5334/gjgl.869

Costamagna, L. (2010). L'apprendimento della fonologia dell'italiano da parte di studenti sinofoni: Criticità e strategie [Learning the phonology of Italian by synophonic students: Criticalities and strategies]. In E. Bonvino & S. Rastelli (Eds.), *La didattica dell'italiano a studenti cinesi e il progetto Marco Polo, Atti del XV seminario AICLU* (pp. 49–66). Pavia University Press.

Costamagna, L., Montilli, C., & Ricci, I. (2014). The role of prosodic features in the acquisition of consonant gemination by Chinese learners. In L. Costamagna & C. Celata (Eds.), *Consonant Gemination in First and Second Language Acquisition* (pp. 47–82). Pacini Editore SpA.

D'Annunzio, B. (2009). *Lo studente di origine cinese. Risorse per docenti di italiano come L2 e LS [The student of Chinese origin. Resources for teachers of Italian as L2 and LS]*. Guerra Edizioni.

Dalla Palma, A. (2019). *Italiano L2 di cinesi. Un caso di tre fratelli sinofoni [Italian L2 of Chinese. A case of three synophonic brothers]* [Unpublished master's thesis]. Università degli Studi di Padova.

Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63–72. https://doi.org/10.1016/j.lindif.2015.04.005

de Leeuw, E., Stockall, L., Lazaridou-Chatzigoga, D., & Gorba Masip, C. (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*(4), 587–618. https://doi.org/10.1177/0267658319886623

De Meo, A., & Petorino, M. (2010). L'acquisizione della competenza prosodica in italiano L2 da parte di studenti sinofoni. In E. Bonvino & S. Rastelli (Eds.), *La didattica dell'italiano a studenti cinesi e il progetto Marco Polo, Atti del XV seminario AICLU* (pp. 67–78). Pavia University Press.

De Meo, A., Vitale, M., & Xu, Y. (2015). Lo sviluppo del sistema vocalico dell'italiano nell'interlingua di adulti sinofoni. In M. Vayra, C. Avesani, & F. Tamburini (Eds.), *Il farsi e il disfarsi del linguaggio: Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio* (pp. 191–202). Offcinaventuno.

Di Benedetto, M. G., & De Nardis, L. (2021a). Consonant gemination in Italian: The affricate and fricative case. *Speech Communication*, *134*, 86–108. https://doi.org/10.1016/j.specom.2021.07.005

Di Benedetto, M. G., & De Nardis, L. (2021b). Consonant gemination in Italian: The nasal and liquid case. *Speech Communication*, *133*, 62–80. https://doi.org/10.1016/j.specom.2021.07.006

Ding, H., Zhan, Y., Liao, S., & Yuan, J. (2019). Production of English stops by Mandarin Chinese learners. *Proceedings of the 9th International Conference on Speech Prosody*, 888–892.

Dmitrieva, O. (2018). The role of perception in the typology of geminate consonants: Effects of manner of articulation, segmental environment, position, and stress. *Language and Speech*, *61*(1), 43–70. https://doi.org/10.1177/0023830917696113

Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (Melopsittacus undulatus): The voiced-voiceless distinction. *Perception & Psychophysics*, *46*(1), 65–71.

Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

Eckman, F. R., Iverson, G. K., & Song, J. Y. (2013). The role of hypercorrection in the acquisition of L2 phonemic contrasts. *Second Language Research*, *29*(3), 257–283. https://doi.org/10.1177/0267658312467029

Eilers, R. E., Wilson, W. R., & Moore, J. M. (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), 1–18.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Netherlands Graduate School of Linguistics.

Esposito, A. (2002). On Vowel Height and Consonantal Voicing Effects: Data from Italian. *Phonetica*, *59*(4), 197–231. https://doi.org/10.1159/000068347

Esposito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America*, *106*(4), 2051–2062. https://doi.org/10.1121/1.428056

Falahati, R. (2015). The production of Persian rhotics by native Mandarin speakers. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1–5). Glasgow, UK: the University of Glasgow.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

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

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217–244). Wissenschaftlicher Verlag.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*(4), 155–171.

Flege, J. E., & Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202.

Flege, J. E., & Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83.

Flege, J. E., & 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*(2), 729–740.

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression* (Third).

Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Francis, A. L., Ciocca, V., & Ching Yu, J. M. (2003). Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America*, *113*(2), 1025–1032.

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

Gamer, M., Lemon, J., & Singh, I. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. https://CRAN.R-project.org/package=irr

Gamkrelidze, T. V. (1975). On the correlation of stops and fricatives in a phonological system. *Lingua*, *35*(3–4), 231–261. https://doi.org/10.1016/0024-3841(75)90060-1

Golestani, N., & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cerebral Cortex*, *17*(4), 929–934. https://doi.org/10.1093/cercor/bhl003

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*(3), 575–582. https://doi.org/10.1093/cercor/bhk001

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech*, *31*(1), 57–75. https://doi.org/10.1177/002383098803100103

Grosjean, F. (2007). The bilingual's language modes. In W. Li (Ed.), *The bilingualism reader* (Second, pp. 428–449). Routledge.

Hall, T. A., & Hamann, S. (2010). On the cross-linguistic avoidance of rhotic plus high front vocoid sequences. *Lingua*, *120*(7), 1821–1844. https://doi.org/10.1016/j.lingua.2009.11.004

Hamzah, H. (2010). Durational properties of initial geminate consonants in Kelantan Malay. *Proc. 13th Australasian International Conference on SST*, 18–21.

Han, M. S. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica*, *49*(2), 102–127. https://doi.org/10.1159/000261906

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *The Journal of the Acoustical Society of America*, *96*(1), 73–82. https://doi.org/10.1121/1.410376

Harada, T. (2006). The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition*, *28*(04). https://doi.org/10.1017/S0272263106060281

Hassan, Z. M. (2002). Gemination in Swedish and Arabic with a particular reference to the preceding vowel duration: An instrumental and comparative approach. *Proceedings of Fonetik*, *44*, 81–84.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: An investigation of English /r/-/l/ perception and production by adult Japanese speakers. In M. Nacano (Ed.), *Interspeech Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*. Waseda University.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5–33. https://doi.org/10.1177/0267658307082980

Hayes, R. L. (2002). The Perception of Novel Phoneme Contrasts in a Second Language: A Developmental Study of Native Speakers of English Learning Japanese Singleton and Geminate Consonant Contrasts. *Coyote Papers*, *12*, 28–41.

Hirata, Y. (2004). Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts. *Computer Assisted Language Learning*, *17*(3–4), 357–376. https://doi.org/10.1080/0958822042000319629

Hirata, Y. (2017). Second language learners' production of geminate consonants in Japanese. In H. Kubozono (Ed.), *The Phonetics and Phonology of Geminate Consonants* (pp. 163–183). Oxford University Press.

Hu, W. (2020). A panel study of Japanese plosive learning by Chinese learners: Taking the percetpion of pa and ba as example. *Nihougo No Gakushu to Kenkyu*, *1*, 61–70. https://doi.org/10.13508/j.cnki.jsr.2020.01.012

Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology*, *2*(2), 301–329. https://doi.org/10.1515/labphon.2011.011

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association*, *38*(2), 167–186. https://doi.org/10.1017/S0025100308003459

Inceoglu, S. (2019). Individual Differences in L2 Speech Perception: The Role of Phonological Memory and Lipreading Ability. *The Modern Language Journal*, *103*(4), 782–799. https://doi.org/10.1111/modl.12591

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*(07), 74–76.

Johnson, K. (2012). *Acoustic and auditory phonetics*. Wiley-Blackwell.

Jusczyk, P. W., Rosner, B. S., Reed, M. A., & Kennedy, L. J. (1989). Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts? *The Journal of the Acoustical Society of America*, *85*(4), 1741–1749.

Kabak, B., Reckziegel, T., & Braun, B. (2011). Timing of second language singletons and geminates. In W.-S. Lee & E. Lee (Eds.), *Proceedings of the 17th International Congress of the Phonetic Sciences (ICPhS XVII)* (pp. 994–997). City University of Hong Kong.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1246. https://doi.org/10.3389/fpsyg.2014.01246

Kassaian, Z. (2011). Age and gender effect in phonetic perception and production. *Journal of Language Teaching and Research*, *2*(2), 370–376. doi:10.4304/jltr.2.2.370-376

Kawahara, S. (2015). The phonetics of sokuon, or geminate obstruents. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 43–78). De Gruyter Mouton.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 286–319.

Keating, P. A., Mikoś, M. J., & Ganong III, W. F. (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*(5), 1261–1271.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221. https://doi.org/10.1121/1.380986

Kluge, D. C., Rauber, A. S., Reis, M. S., & Bion, R. A. H. (2007). The relationship between the perception and production of English nasal codas by Brazilian learners of English. *Proceedings Interspeech 8*, 2297–2300.

Kochetov, A. (2005). Phonetic sources of phonological asymmetries: Russian laterals and rhotics. *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*, 1–12.

Kramer, M. (2009). *The phonology of Italian*. Oxford University Press on Demand.

Kubozono, H. (2013). Introduction to the special issue on Japanese geminate obstruents. *Journal of East Asian Linguistics*, *22*(4), 303–306.

Kubozono, H. (2017). Introduction to the phonetics and phonology of geminate consonants. In H. Kubozono (Ed.), The phonetics and phonology of geminate consonants (pp. 1–10). Oxford University Press.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69–72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*(3), 905–917.

Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, *5*(2), 13–24.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.

Lee, A., & Mok, P. (2018). Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research*, *34*(4), 419–448. https://doi.org/10.1177/0267658317739056

Lee, S. J., Kurita, G., & Perkins, J. (2018). Identifying the locus of L2 pronunciation: An exploratory study of geminate production by Chinese L2 learners of Japanese. *Foreign Language Education Research*, *22*, 69–86.

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112. https://doi.org/10.1017/S0025100303001208

Lee, W.-S., & Zee, E. (2008). Prosodic characteristics of the neutral tone in Beijing Mandarin. *Journal of Chinese Linguistics*, *36*(1), 1–29.

Lehiste, I. (1970). *Suprasegmentals*. Massachusetts Inst. of Technology P.

Lennes, M. (2002). *Calculate_segment_durations.praat*. https://github.com/lennes/spect/blob/1.0.0/scripts/calculate_segment_durations.praat

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.5.3) [Computer software]. https://CRAN.R-project.org/package=emmeans

Li, P., Baills, F., & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Studies in Second Language Acquisition*, *42*(5), 1015–1039. https://doi.org/10.1017/S0272263120000054

Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi [Acoustic characteristics of neutral tone in Beijing Mandarin]. *Fangyan*, *3*, 166–178.

Lin, T. (1985). Tantao Beijinghua qingyin xingzhi de chubu shiyan [On neutral tone in Beijing Mandarin]. In S. Hu (Ed.), *Beijing yuyin shiyanlu [Working papers in experimental phonetics]* (pp. 1–26). Beijing Daxue chubanshe [Peking University Press].

Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge University Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422.

Liu, J., Zeng, T., & 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.

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

Luo, J., Li, V. G., & Mok, P. P. K. (2020). The Perception of Cantonese Vowel Length Contrast by Mandarin Speakers. *Language and Speech*, *63*(3), 635–659. https://doi.org/10.1177/0023830919879471

Maddieson, I. (1985). Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 203–221). Academic Press.

Mairano, P., & Calabrò, L. (2017). Are minimal pairs too few to be used in pronunciation classes? In R. Savy & I. Alfano (Eds.), *La fonetica sperimentale nell''insegnamento e nell''apprendimento delle lingue straniere, collana Studi AISV: Vol. II* (pp. 255–268). Officinaventuno.

Mairano, P., & De Iacovo, V. (2020). Gemination in Northern versus Central and Southern Varieties of Italian: A Corpus-based Investigation. *Language and Speech*, *63*(3), 608–634. https://doi.org/10.1177/0023830919875481

Marotta, G., & Vanelli, L. (2021). *Fonologia e prosodia dell'italiano [Italian phonology and prosody]*. Carocci.

Martinet, A. (1952). Function, structure, and sound change. *Word*, *8*(1), 1–32.

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, *33*(2), 255–259. https://doi.org/10.1017/S0025100303001373

Masuko, Y., & Kiritani, S. (1990). Perception of Mora Sounds in Japanese by Non-Native Speakers of Japanese. In Acoustical Society of Japan (Ed.), *Proceedings: 1990 International Conference on Spoken Language Processing* (pp. 1333–1336). Acoustical Society of Japan.

Meng, Y., Kotzor, S., Xu, C., Wynne, H. S., & Lahiri, A. (2021). Mismatch negativity (MMN) as an index of asymmetric processing of consonant duration in fake Mandarin geminates. *Neuropsychologia*, *163*, 1–8. https://doi.org/10.1016/j.neuropsychologia.2021.108063

Minagawa-Kawai, Y., & Kiritani, S. (1998). Non-native productions of Japanese single stops that are too long for one mora unit. *Proceedings of the 5th International Conference on Spoken Language Processing*.

Minagawa, Y., & Kiritani, S. (1996). Discrimination of the single and geminate stop contrast in Japanese by five different language groups. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, *30*, 23–28.

Motohashi-Siago, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, *13*(2), 29–47.

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

Nearey, T. M., & Rochet, B. L. (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), 1–18. https://doi.org/10.1017/S0025100300004965

Neufeld, G. G. (1979). Towards a theory of language learning ability. *Language Learning*, *29*(2), 227–241. https://doi.org/10.1111/j.1467-1770.1979.tb01066.x

Nodari, R., Celata, C., & 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. https://doi.org/10.1016/j.wocn.2018.12.005

Ortí Mateu, R. (1990). *Comparación fonética, diagnóstico y tratamiento de las dificultades de los estudiantes chinos para aprender español [Phonetic comparison, diagnosis and treatment of Chinese students' difficulties in learning Spanish]* [Unpublished doctoral dissertation]. University of the Philippines.

Patience, M. (2018). Acquisition of the tap-trill contrast by L1 Mandarin–L2 English–L3 Spanish Speakers. *Languages*, *3*(4), 42. https://doi.org/10.3390/languages3040042

Payne, E. M. (2005). Phonetic variation in Italian consonant gemination. *Journal of the International Phonetic Association*, *35*(2), 153–181. https://doi.org/10.1017/S0025100305002240

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. *Proceedings Interspeech 12*, 161–164.

Peterson, R. A., & Cavanaugh, J. E. (2020). Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, *47*(13–15), 2312–2327. https://doi.org/10.1080/02664763.2019.1630372

Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of Speaking Rate on the Singleton/Geminate Consonant Contrast in Italian. *Phonetica*, *56*(3–4), 135–157. https://doi.org/10.1159/000028448

R Core Team. (2020). *R: A language and environment for statistical computing*. [Computer software]. https://www.R-project.org/

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., & Dupoux, E. (2010). A psycholinguistic perspective on the acquisition of phonology. *Laboratory Phonology*, *10*(3), 311–340. https://doi.org/10.1515/9783110224917

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, *35*(3), 439–443. https://doi.org/10.1016/j.wocn.2006.01.002

Ringen, C., & Kulikov, V. (2012). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics*, 269–286.

Rochet, B. L., & Chen, F. (1992). Acquisition of the French VOT contrasts by adult speakers of Mandarin Chinese. *Second International Conference on Spoken Language Processing Proceedings*, 273–276.

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

Rogers, D., & d'Arcangeli, L. (2004). Italian. *Journal of the International Phonetic Association*, *34*(1), 117–121. https://doi.org/10.1017/S0025100304001628

Rojczyk, A. (2011). Perception of the English voice onset time continuum by Polish learners. In J. Arabski & A. Wojtaszek (Eds.), *The Acquisition of L2 Phonology* (pp. 37–58). Multilingual Matters.

Romano, A. (2013). A preliminary contribution to the study of phonetic variation of /r/ in Italian and Italo-Romance. In L. Spreafico & A. Vietti (Eds.), *Rhotics. New Data and Perspectives* (1st ed., pp. 209–225). Bozen-Bolzano University Press.

Rossetti, R. (1994). Gemination of Italian stops. *The Journal of the Acoustical Society of America*, *95*(5), 2874–2874. https://doi.org/10.1121/1.409450

Serniclaes, W. (2005). On the invariance of speech percepts. *ZAS Papers in Linguistics*, *40*, 177–194.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

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

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

Sun, P. P., & Zhang, L. J. (2020). A Multidimensional Perspective on Individual Differences in Multilingual Learners' L2 Chinese Speech Production. *Frontiers in Psychology*, *11*.

Sun, Y., & Profita, S. (2020). Cross-Linguistic Study on VOT of Chinese Trilingual Speakers. *Studies in Literature and Language*, *20*(1), 98–102.
Sun, Y., & Profita, S. (2020). Cross-Linguistic Study on VOT of Chinese Trilingual Speakers. *Studies in Literature and Language*, *20*(1), 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*, 171–174.

Tanaka, S. (2017). The relation between L2 perception and L1 phonology in Japanese loanwords: An analysis of geminates in loanwords from Italian. In H. Kubozono (Ed.), *The Phonetics and Phonology of Geminate Consonants* (pp. 321–339). Oxford University Press.

Toda, T. (1997). Strategies for producing mora timing by non-native speakers of Japanese. In T. Toda (Ed.), *Acquisition of Japanese as a Second Language.* (pp. 157–197). John Benjamins Publishing.

Toda, T. (2003). Acquisition of special morae in Japanese as a second language. *Journal of the Phonetic Society of Japan*, *7*(2), 70–83.

Trifone, P. (2014). L'italiano nella giostra delle lingue. Testi di apprendenti sinofoni, ispanofoni e slavofoni [Italian in the carousel of languages. Texts by Sinophonic, Hispanic and Slavophonic learners.]. In P. Danler & C. Konecny (Eds.), *Dall'architettura della lingua italiana all'architettura delle lingue. Saggi in omaggio a Heidi-Siller Runggaldier* (pp. 639–647).

Tsukada, K., & Hajek, J. (2019a). Cross-language perception of Italian and Japanese consonant length contrasts: A comparison of native Italian listeners with and without Japanese language learning experience. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 82–86). Australasian Speech Science and Technology Association Inc.

Tsukada, K., & Hajek, J. (2019b). Misperception of Italian Singleton and Geminate Obstruents by Native Japanese Speakers. *Journal of the Phonetic Society of Japan*, *23*, 148–156.

Tsukada, K., & Hajek, J. (2020). Perception of consonant length in familiar and unfamiliar languages by native speakers of Mandarin, Italian and Japanese. *Proceedings of the 10th International Conference on Speech Prosody* (pp. 66–70). https://doi.org/10.21437/SpeechProsody.2020-14

Tsukada, K., Cox, F., Hajek, J., & Hirata, Y. (2018). Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Second Language Research*, *34*(2), 179–200. https://doi.org/10.1177/0267658317719494

Tsukada, K., Hirata, Y., & Roengpitya, R. (2014). Cross-Language Perception of Japanese Vowel Length Contrasts: Comparison of Listeners From Different First Language Backgrounds. *Journal of Speech, Language, and Hearing Research*, *57*(3), 805–814. https://doi.org/10.1044/2014_JSLHR-S-12-0416

Vagges, K., Ferrero, F. E., Magno-Caldognetto, E., & Lavagnoli, C. (1978). Some acoustic characteristics of Italian consonants. *Journal of Italian Linguistics Amsterdam*, *3*(1), 69–84.

Vale, A. (2020). *Perceção das Consoantes líquidas / ɾ / e / l / do Português Europeu sob Influência do Mandarim L1 [Perception of the liquid consonants /ɾ/ and /l/ of European Portuguese under the Influence of Mandarin L1]* [Unpublished master's thesis]. University of Minho.

Valentini, A. (1990). Lo palale italiano no bene. *Italiano e Oltre*, *1*, 37–39.

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852–866.

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).

Yamakawa, K., Amano, S., & Kondo, M. (2021). Mispronunciation of Japanese singleton and geminate stops by Korean and Taiwanese Mandarin speakers. *Acoustical Science and Technology*, *42*(2), 73–82. https://doi.org/10.1250/ast.42.73

Yang, Y., & Chen, X. (2019). Within-organ contrast in second language perception: The perception of Russian initial /r-l/ contrast by Chinese learners. *The Journal of the Acoustical Society of America*, *146*(2), EL117–EL123. https://doi.org/10.1121/1.5120549

Yang, Y., & Fang, Z. (1984). Putonghua songqi he busongqi seyin de yinwei jiexian jiqi fanchou zhijue [Phonemic boundary and categorical perception of Mandarin

Chinese stops]. *Quanguo Diwujie Xinlixue Xueshu Huiyi Wenzhai Xuanji [Proceedings of the 5th Chinese National Psychology Conference]*, 220–229.

Yang, Y., Chen, X., & 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*(1), 3–29. https://doi.org/10.1177/0267658319900919

Yanushevskaya, I., & Bunčić, D. (2015). Russian. *Journal of the International Phonetic Association*, *45*(2), 221–228. https://doi.org/10.1017/S0025100314000395

Yuan, J. (2001). *Hanyu fanyan gaiyao [Outline of Chinese dialects]*. 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*, 174–181. https://doi.org/10.14095/b.cnki.jics.2012.02.028

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*(03), 58–64.

Zhang, S. (2013). Zhongguo xuesheng fayu xuexizhong 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*(8), 100–102.

Zhou, C. (2021). *L2 speech learning of European Portuguese /l/ and /ɾ/ by L1-Mandarin learners: Experimental evidence and theoretical modelling* [Unpublished doctoral dissertation]. Universidade de Lisboa.

Zhou, C., Freitas, M. J., & Castelo, A. (2021). On the acquisition of European Portuguese liquid consonants by L1-Mandarin learners. In J. Pinto & A. Nélia (Eds.), *Multilingualism and third language acquisition: Learning and teaching trends.* (pp. 25–45). Language Science Press.

Zhou, T. (2012). Analysis of common pronunciation errors made by Chinese students in Italian learning and some suggestions to Italian pronunciation teaching. *Chinese Educational Technology & Equipment*, *33*, 98–99.

Appendix

I. Word stimuli for the production experiment of Chapter 2.

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>

II. Word stimuli for the perception and production (in bold) experiments of Chapter 3.

Target words	Distractors
papa 'pope' vs. ***pappa*** 'pulp'	***passo*** vs. ***basso***
dita 'fingers' vs. ***ditta*** 'company'	***pollo*** vs. ***bollo***
Luca 'person's name' vs. ***Lucca*** 'city's name'	***tetto*** vs. ***detto***
pena 'penalty' vs. ***penna*** 'pen'	***quanto*** vs. ***quando***
pala 'shovel' vs. ***palla*** 'ball'	***cara*** vs. ***gara***
cade 'falls' vs. *cadde* 'fell'	*eco* vs. *ego*
face 'torch' vs. *facce* 'faces'	*pasta* vs. *basta*
bevi 'drink' vs. *bevvi* 'drank'	*panca* vs. *banca*
agio 'ease' vs. *aggio* 'premium'	*pari* vs. *Bari*
tufo 'tuff' vs. t*uffo* 'dip'	*noto* vs. *nodo*

III. Word stimuli for the perception and production experiments of Chapter 4.

Target words	Distractors
	passo vs. *basso*
	pollo vs. *bollo*
rana ‘frog’ vs. lana ‘wool’	*tetto* vs. *detto*
regno ‘kingdom’ vs. *legno* ‘wood’	*quanto* vs. *quando*
mare ‘sea’ vs. *male* ‘badly’	*cara* vs. *gara*
cara ‘dear’ vs. *cala* ‘drops’	*eco* vs. *ego*
farsi ‘make oneself’ vs. *falsi* ‘fake’	*pasta* vs. *basta*
Parma ‘city name’ vs. *palma* ‘palm’	*panca* vs. *banca*
	pari vs. *Bari*
	noto vs. *nodo*