


Article

The Acquisition of Mandarin Consonants by English Learners: The Relationship between Perception and Production

Xinchun Wang * and Jidong Chen 

Department of Linguistics, College of Arts and Humanities, California State University, Fresno, CA 93740, USA; jchen@csufresno.edu

* Correspondence: xinw@csufresno.edu

Received: 15 February 2020; Accepted: 9 May 2020; Published: 13 May 2020



Abstract: This study investigates native English CFL (Chinese as a Foreign Language) learners' difficulties with Mandarin consonants at the initial stage of learning and explores the relationship between second language (L2) speech perception and production. Twenty-five native English CFL learners read the eight Mandarin consonants (j/ɟ/, q/ɟʰ/, x/ç/, zh/ʈʂ/, ch/ʈʂʰ/, sh/ʃ/, z/ʈʂ/, and c/ʈʂʰ/) in sentences and identified the target sounds in a forced-choice identification task. Native Mandarin listeners identified the consonants produced by the learners and rated the quality of each sound they identified along a scale of 1 (poor) to 7 (good). The learners' mean percentage accuracy scores ranged from 29% to 80% for perception and 25% to 88% for production. Moderate correlations between the perception and production scores were found for two of the eight target sounds. The Mandarin retroflex, palatal, and dental fricatives and affricates, though all lack counterparts in English, pose different problems to the English CFL learners. The misperceived retroflex and palatal sounds were substituted with each other in perception but mis-produced palatal sounds were substituted with each other, not with retroflex sounds. The relationship between perception and production of L2 consonants is not straightforward. The findings are discussed in terms of current speech learning models.

Keywords: Mandarin consonants; English CFL learners; L2 perception and production

1. Introduction

1.1. L2 Speech Perception and Perception Models

Adult second language (L2) speakers' problems with the perception and production of non-native speech sounds are closely related to their first language (L1) experience (Flege 1995). Research has shown that infants are language-general perceivers of speech sounds at the phonetic level. This universal perceptual pattern undergoes a profound change due to increased experience with their first language in the later half of the first year in life (Best 1994; Polka and Bohn 1996; Strange 1995; Werker 1994; Werker and Polka 1993). Adult monolinguals are language-specific perceivers of speech sounds. Perceptual studies using synthesized stimuli found that adult speakers identified stop consonants along a VOT (Voice Onset Time) continuum according to their L1 stop inventories (Lisker and Abramson 1964, 1970). Similar studies on L2 vowel perception using a synthesized vowel continuum also indicated that listeners labeled vowel sounds according to their L1 vowel categories (Rochet 1995). To a large extent, the language-specific nature of adult monolinguals' speech perception underlies the difficulties adult learners face in L2 speech learning.

Evidence from cross-linguistic speech perception studies in which listeners map L2 sounds onto their L1 sound system suggest that the phonetic distances between learners' L1 and L2 sound systems

play an important role in the degree of success in L2 speech perception (Flege and Wayland 2019; Guion et al. 2000; Wang and Chen 2019). For example, in a cross-linguistic perceptual study assessing the phonetic distances between Japanese and English consonants, monolingual Japanese speakers identified English consonants using Japanese consonant categories. The subsequent experiment found that phonetic distances between Japanese and English consonants, as established by the cross-linguistic direct mapping experiment, predicted the discrimination patterns of English consonants by Japanese learners of English with different L2 experience (Guion et al. 2000). Wang and Chen (2019) also found that English CFL learners' perception problems with Mandarin consonants were closely related to the L2 to L1 assimilation patterns.

In searching for the nature of such cross linguistic influence in L2 phonetic learning, researchers have come up with different L2 speech perception models. The Perceptual Assimilation Model (PAM) (Best 1994; Best et al. 2001) assumes that several pairwise assimilation types are possible when two non-native phones are mapped onto the L2 sound system. The pair of L2 phones may be assimilated to two different L1 phones, the Two Category (TC) type, or to a single L1 category equally poorly or well, the Single Category type (SC). The two L2 sounds can also be assimilated to a single native category but one can be a better fit than the other, the Category Goodness type (CG). The PAM model also predicts the degree of difficulties in discriminations of L2 sounds from the most to the least: SC > CG > TC (Best et al. 2001).

Flege (1995, 2007) Speech Learning Model (SLM) states that a learner's L1 and L2 sound systems interact and exist in a common phonological space. Learners will establish an L2 sound category if they perceive the phonetic differences between the L2 sound from the nearest L1 sound or the closest L2 sound. In contrast, "equivalence classification" of an L2 sound with the nearest L1 category blocks the formation of a new phonetic category. Flege claims that learners' ability to establish new phonetic categories remains intact throughout their life span and increases with their L2 experience. Perceptual learning will eventually lead to better production, although the alignment between perception and product may be partial only (Flege 1999). Therefore, the SLM is a dynamic model that emphasizes learners' L2 experiences with the target language.

1.2. The Relationship between Perception and Production

As both the PAM and SLM models place more emphasis on the perceptual assimilation or dissimilation of the L2 sound categories to the L1 sounds, the question arises about the relationship between L2 speech perception and production. Previous research on L2 speech perception and production has led to different conclusions. For example, Rochet (1995) found that native Portuguese speakers produced French /y/ as /i/ while native English speakers produced /y/ as /u/, although both English and Portuguese have /i/ and /u/ in their vowel systems. The subsequent perceptual test using synthesized high vowel continuum revealed that Portuguese listeners assimilated /y/ to /i/ while English listeners assimilated /y/ to /u/ (Rochet 1995). Similarly, Mandarin speakers' production problem with French voiced stops was related to their faulty perception of the voiced stops that do not exist in the Mandarin sound system (Rochet 1995). In a study on English front vowels /i I EI εæ/, Wang (1997) found that Mandarin speakers had problems with both the perception and production of English lax vowels /I εæ/, but they performed better in perception than in production on these three vowels. In contrast, they performed better in production than in perception on English /i EI/ categories. Such performance discrepancies between the perception and production on the English front vowels suggest that native Mandarin ESL (English as a Second Language) learners may have used different cues or strategies in their perception and production of English vowels. Flege (1999) also reported a series of studies that showed partial alignment between L2 perception and production. In a more recent study on native Arabic speakers' acquisition of British English vowels and consonants, Evans and Alshangiti (2018) found a link between perception and production, as the better perceivers of English vowels were also the better producers.

L2 phonetic training studies have also examined the relationship between perception and production when assessing the effects of training in both modes. In a recent review study applying the meta-analysis method analyzing 30 perception training studies on L2 segments conducted in the past 25 years, Sakai and Moorman (2018) found that perception training only led to small-sized gains in productions of the target sounds. Their subsequent statistical analysis based on 18 out of the 30 studies led to the conclusion that the production gains were larger on obstruents than on sonorants and vowels. Correlation tests suggested there was a small to medium-sized but statistically nonsignificant relationship between gains in perception and production.

1.3. Studies on L2 Mandarin Consonants

While L2 Learners' problems with non-native speech sounds are well documented on consonants (Bradlow et al. 1997; Guion et al. 2000; Munro et al. 2015), and on vowels (Evans and Alshangiti 2018; Munro and Derwing 2008; Wang 1997; Wang and Munro 2004), as well as on lexical tones (Wang 2006, 2008, 2013), parallel studies on perception and production of L2 speech sounds, particularly on CFL learners' difficulties with Mandarin consonants are still very limited. Several studies on the perception or production of Mandarin consonants reported in the past two decades are summarized in the following.

Lai (2009) investigated learners' perception difficulties with the six Mandarin affricates z /ts/, c /ts^h/, zh /tʂ/, ch /tʂ^h/ and j /tʃ/, and q /tʃ^h/ by native Malay and Burmese speakers residing in Taiwan. The learners and a control group of native Taiwan Mandarin speakers took the same/different discrimination test followed immediately by the identification test on the target affricates paired across different place and manner of articulations. Both learner groups were more accurate in identifying unaspirated affricates than the aspirated counterparts. They also had more problems identifying the dental-retroflex z /ts/- zh /tʂ/ and c /ts^h/- ch /tʂ^h/ contrasts than the palatal affricates. Lai (2009) concluded that there was a merge of dental and retroflex affricates and the dentalization of the retroflex sounds was better explained by the Markedness theory than the learners' first language inference. In fact, the native Mandarin control group demonstrated exactly the same perceptual merge pattern as they had the same rate of errors (around 67%) as the two learner groups on their z /ts/- zh /tʂ/ and c /ts^h/- ch /tʂ^h/ identifications. The findings were not surprising as both L2 groups were learning Mandarin in Taiwan and the dental and retroflex fricative/affricate merge is common in many Mandarin dialects spoken in Southern China as well as in Taiwan (Zhu 2012; Chuang et al. 2019).

Hao (2012) investigated how the learners' L2 to L1 sound mapping patterns and the amount of L2 experience affect the perception of Mandarin sounds. Three groups of native English CFL learners with different length of Mandarin learning experience: Ex group (5.6 years), Inex group (1.5 years) and Noex (No experience) took the perceptual tests. Hao (2012) found that phonetic context and L2 Mandarin experience affect the learners' L2 to L1 sound mapping patterns. More experienced learners gave more consistent responses in Mandarin to English sound classifications and were less affected by phonetic contexts than less experienced learners. The Noex group assimilated Mandarin /s/ to English /z/ more often than to /s/ while both learner groups identified /s/ as English /s/. All three groups assimilated Mandarin /ʃ/ and /ç/ to English /f/ mostly except that the Noex group split the classification of /ç/ to /s/ and /f/ equally when /ç/ was followed by an unrounded vowel /i/. While both Mandarin /ʃ/ and /ç/ were assimilated to the English /f/, /ʃ/ was a better fit than /ç/ as indicated by both the identification accuracy rate and the higher goodness rating score, a Category Goodness type of assimilation according to the PAM model. In the identification test, the Mandarin /ʃ-ç/ contrast was found difficult for the learners and more so for the Inex group than the Ex group. All three groups performed equally well in discriminating the /ʃ u-su/ and /ʃi-si/ contrasts in the discrimination test. The author concluded that the L2 to L1 assimilation patterns failed to predict discrimination accuracy of Mandarin contrasts in most cases.

In a more recent cross-linguistic perception study on Mandarin consonants (Wang and Chen 2019), native English listeners with no Mandarin learning experience identified 10 Mandarin consonants

in syllables (z /tʂa/, c /tʂʰa/, s /sa/, j /tɕʰa/, q /tɕʰa/, x /ɕʰa/, zh /tʂʰa/, ch /tʂʰa/, sh /ʂa/, and r /zɑ/) using the closest English sounds in a ten-way forced choice task followed by a goodness rating task along a scale of 1 (poor) to 7 (good). L2 to L1 sound mapping fitting indexes (identification score × rating score) were calculated to assess the phonetic distances between Mandarin and English consonants. Wang and Chen (2019) found there was a range of phonetic distances between the L1 and L2 sounds based on the fit indexes (range from 1.0 to 6.3 out of 7). The “poor” matching categories were x /ɕ/, c /tʂʰ/, q /tɕʰ/, zh /tʂ/, and j /tɕ/ whose fit indexes were below the mean (3.7, s.d.=1.7). The “fair” fitting categories were ch /tʂʰ/, s /s/, and z /ts/ whose fit indexes were at the mean. The “good” matching sounds were r /z/, and sh /ʂ/ whose fit indexes were 1s.d. above the mean (Wang and Chen 2019). In a subsequent study on the identification of Mandarin consonants by English CFL learners at two different proficiency levels, the learners’ perception scores of Mandarin consonants were found to be closely related to the L2 to L1 assimilation patterns. Results showed that zh /tʂ/, q /tɕʰ/, c /tʂʰ/, and x /ɕ/ (the poor fitting sounds) received the lowest % identification scores among the 10 sounds by the beginning level learners. The intermediate group outperformed the beginning group on zh /tʂ/, q /tɕʰ/, and c /tʂʰ/. These findings suggest that the perceived phonetic distances between L1 and L2 consonants predicted the English CFL learners’ L2 Mandarin consonant identification problems and increased L2 experience improved perceptual learning. No production data were reported in this study.

In a production study, Liu and Jongman (2012) investigated both the temporal and spectral features of Mandarin dental affricates z /ts/, and c /tʂʰ/ produced by native English CFL learners with different proficiency levels. The authors found that both the novice and more experienced learner groups acquired the durational differences for the /ts/, and /tʂʰ/ contrast but only the more advanced learners acquired the spectral (center of gravity) contrast between the target sound pair. It was not clear what weight the temporal and spectral cue each carries to the perceptual accuracy of the target contrast as no perception test was conducted to measure the accuracy of the learners’ productions. This study dealt with only one pair of Mandarin affricate contrast at dental place of articulation.

In a similar study involving more Mandarin affricate contrasts, Yang and Yu (2019) investigated the perception and production of six Mandarin affricates z /ts/, c /tʂʰ/, zh /tʂ/, ch /tʂʰ/, j /tɕ/, and q /tɕʰ/ by native English CFL learners at beginning and intermediate levels. Both learner groups matched the native Mandarin group in perception accuracy scores in discriminating but not in identifying the target sounds. The effect of place of articulation and aspiration were significant but not uniform across the board. For example, the unaspirated palatal j /tɕ/ was significantly better identified than the aspirated palatal counterpart q /tɕʰ/ but the aspirated retroflex ch /tʂʰ/ was better identified than the unaspirated counterpart zh /tʂ/. The authors concluded that different affricates pose different learning difficulties for English CFL learners. In the production test, the intermediate group outperformed the beginning group in approximating the native speakers in the production of some but not all the acoustical features under investigation, indicating the learners did not acquire the affricates completely. Their data suggest that the distinction between palatal and retroflex affricates is more difficult for learners due to the assimilation of both classes to the same English post-alveolar affricates, the two to one type of (SC) of assimilation, according to the PAM model.

To summarize the findings of the above studies, the difficulties with the perception accuracy of Mandarin consonants by native English CFL learners are related to their L2 to L1 perceptual assimilation patterns (Hao 2012; Wang and Chen 2019; Yang and Yu 2019). In general, Mandarin retroflex and palatal contrasts pose more difficulties to the English CFL learners than other place contrasts (Hao 2012; Yang and Yu 2019), while dental and retroflex contrasts were more difficult for Malay and Burmese learners (Lai 2009). The effect of L2 experience did not appear to affect the learners’ discrimination accuracy but did influence their identification accuracy of the Mandarin consonants (Hao 2012; Lai 2009; Wang and Chen 2019; Yang and Yu 2019). These results confirmed earlier findings of the advantage of identification over discrimination task in L2 phonetic test and training because the former help the learners focus more on the key phonetic/acoustic features that distinguish the target sound contrasts (Wang and Munro 2004). In production, the effect of L2 experience was more evident

as the more experienced learners outperformed less experienced learners in approximating the native speakers in the production of some but not all the acoustical features of the target sound contrasts (Liu and Jongman 2012; Yang and Yu 2019).

1.4. The Current Study

Several studies summarized in Section 1.3 (Hao 2012; Lai 2009; Wang and Chen 2019) investigated CFL learners’ perception problems with Mandarin consonants but did not examine their production problems. Two production studies on Mandarin consonants (Liu and Jongman 2012; Yang and Yu 2019) compared the acoustic properties of the learners’ productions with those of the native speakers but did not include the direct assessment of the intelligibility of the L2 speech. Parallel studies that compare the CFL learners’ perception and production performance with Mandarin consonants are extremely rare. This study aims to fill this gap by investigating native English CFL learners’ difficulties with the Mandarin consonants in both perception and production at initial stage of learning. An additional goal is to examine the relationship between L2 speech perception and production. The research questions are:

1. Which Mandarin consonants are difficult to identify and produce for native English CFL learners at early stage of learning?
2. How do the phonetic differences and distances between Mandarin and English consonants, as perceived by English listeners in an earlier study (Wang and Chen 2019), affect the perception and production of Mandarin consonants?
3. What are the learners’ performance differences between their perception and production of Mandarin consonants?

1.5. Mandarin Consonants

Table 1 presents the 22 Mandarin consonants in IPA. The sounds in bold are the eight target Mandarin consonants under investigation in the current study: **z /ts/, c /ts^h/, j /tɕ/, q /tɕ^h/, x /ç/, zh /tʂ/, ch /tʂ^h/, sh /ʃ/**. They form the fricative/affricate groups at dental, retroflex, and alveolo-palatal (also commonly referred to as palatal) places reported to be difficult for English CFL learners to acquire as these sounds do not have corresponding counterparts in English (Lin 2005; Wang and Chen 2019).

Table 1. Mandarin Consonants.

	Labial	Dental	Retroflex	Palatal	Velar
Stop	p p ^h	t t ^h			k k ^h
Affricate		ts ts^h	tʂ tʂ^h	tɕ tɕ^h	
Fricative	f	s	ʂ ʐ	ç	x
Nasal	m	n			ŋ
Liquid		l			

2. Experiment 1: Perception of Mandarin Consonants

2.1. Participants

The participants were 25 native English speaking (15 male, 10 female, mean age = 19.6) beginning level CFL learners enrolled in a first semester Chinese course in a public university in the U.S. All participants reported speaking English as their native language. Twenty of them were born and raised in the United States and five were born in foreign countries but moved to the U.S. between the ages of 2 and 5. Some participants reported speaking another language along with English as their first languages. They were four English/Spanish, four English/Hmong, two English/Tagalog, and one English/Vietnamese early bilinguals. At the point of data collection, the participants were about three months into the 16-week semester and all had learned and practiced Chinese consonants by then.

2.2. Material

The perceptual identification test initially included 10 Mandarin consonants in syllables (z /tʂa/, c /tʂʰa/, s /ʂa/, j /tɕʰa/, q /tɕʰa/, x /ɕʰa/, zh /tʂa/, ch /tʂʰa/, sh /ʂa/, and r /ʐa/) that were produced by two native Mandarin Speakers, one male and one female, through a reading task. The target words were produced in a carrier sentence *wo shuo ___ zi* (我说一字). “I say — word”. The recordings were made on a MacBook Pro computer using the Praat software. The target syllables were separated from the sentences using waveform editing, normalized for peak volume, and saved as wave form for presentations. Eight of the 10 sounds (z /ts/, c /tsʰ/, j /tɕ/, q /tɕʰ/, x /ɕ/, zh /tʂ/, ch /tʂʰ/, sh /ʂ/) were analyzed for this perception experiment to pair exactly with the eight target consonants in the production test.

2.3. Procedure

All participants gave their informed consent for inclusion before they participated in this study which was approved by the ethics committee of the researchers’ university. Individual perception identification tasks were carried out in a sound booth on a MacBook Pro computer using Praat ExperimentMFC identification test design. A total of 60 stimuli (10 sounds 2 speakers 3 repetitions) were randomized and presented in a 10-way forced choice task. The labels for choices were the 10 consonants in pinyin (the official Romanized transcription of Mandarin Chinese) displayed on the computer screen during the test. The listeners were instructed to listen carefully for the initial consonant in each stimulus and identify the sound they heard by clicking on the corresponding consonant on the screen. During the test, they could choose to replay each stimulus twice in the case of uncertainty. To familiarize the learners with the task, before the real test began, each participant had a trial session using the stimuli not included for analyses. The software automatically recorded the test data to be exported for analysis.

2.4. Results

Individual participants’ correct identifications of each target consonant were converted to percentage accuracy scores. Their misidentified target sounds were also converted to percentage error rate and were tallied for substitution patterns. The group mean percentage correct identification scores of the eight consonants ranged from 29% zh /tʂ/ to 80% ch /tʂʰ/. To investigate the learners’ perceptual substitution patterns of the misidentified consonants, a confusion matrix was created and is presented along with the percentage correct identification scores in Table 2.

Table 2. Mean percentage correct identification scores (in bold) and confusion matrix of Mandarin consonants by native English CFL (Chinese as a Foreign Language) learners (N = 25).

Target	Identified									
	zh /tʂ/	ch /tʂʰ/	sh /ʂ/	j /tɕ/	q /tɕʰ/	x /ɕ/	z /ts/	c /tsʰ/	s /s/	r /ʐ/
zh /tʂ/	29	22	5	32	4	1	3	2	1	1
ch /tʂʰ/	4	80	4		5		1	3	1	1
sh /ʂ/	1	5	70		1	15		3	5	
j /tɕ/	8	11		61	11	3	4	1		
q /tɕʰ/	5	43	3	3	31	4	3	4	3	1
x /ɕ/	2	3	31	3	5	46		3	5	1
z /ts/	7	1		1		3	56	11	20	
c /tsʰ/	5	11	2		7	1	9	53	11	

A One-Way repeated measures ANOVA was conducted to compare the differences between the perception scores on the consonants (8 levels). There was a significant effect of consonant (Wilk’s Lambda = 0.102, F (7, 25) = 22.526, p = 0.000). The subsequent post hoc Bonferroni tests adjusted for multiple comparisons revealed that a series of pairwise comparisons were significant. The results are presented in Table 3.

Table 3. Pairwise comparisons of differences between the consonants in perception (** $p < 0.01$, * $p < 0.05$).

	c /ts ^h /	ch /tʂ ^h /	j /tɕ/	q /tɕ ^h /	sh /ʃ/	x /ç/	z /ts/	zh /tʂ/
c /ts ^h /	-	*						*
ch /tʂ ^h /	*	-		**		**		**
j /tɕ/			-	**				**
q /tɕ ^h /		**	**	-	**		**	
sh /ʃ/				**	-			**
x /ç/		**				-		
z /ts/				**			-	**
zh /tʂ/	*	**	**		**		**	-

2.5. Discussion

The perception test results showed that Mandarin zh /tʂ/ (29%), q /tɕ^h/ (31%), and x /ç/ (46), were the worst identified sounds by native English CFL learners. The results of the pairwise comparisons confirmed that the perception scores of zh /tʂ/ (29%), q /tɕ^h/ (31%), and x /ç/ (46) were significantly different from all the other five sounds but not different from each other (See Table 3). The findings were similar to the Wang and Chen (2019) study in which the same three sounds were also among the four most difficult consonants for the beginning level learners. These three sounds were also among the poorest fitting categories to the learners’ L1 English sounds as established by the native English listeners in a cross-linguistic identification test found in the Wang and Chen (2019) study. The current findings support the Wang and Chen (2019) findings that phonetic distances and differences between L1 and L2 consonants predicted native English CFL learner’s perception problems with Mandarin consonants at initial stage of learning.

An inspection of the confusion matrix of the misidentified sounds led to the observation that misidentified retroflex and palatal sounds are mostly confused with each other. For example, the retroflex zh /tʂ/ was heard as palatal j /tɕ/ 32% of times. The confusion score exceeded the correct % identifications of zh /tʂ/ of only 29%. Similarly but to a less degree, the highest % of misidentified sh /ʃ/ was heard as x /ç/ 15%. Misidentified palatal sounds were heard mostly as retroflex sounds as well. The palatal sounds q /tɕ^h/ and x /ç/ were misidentified as the retroflex sounds ch /tʂ^h/ and sh /ʃ/ 43% and 31% of times, respectively. Therefore, the retroflex and palatal fricatives and affricates are both difficult for English CFL learners to identify.

The dental affricates z /ts/ and c /ts^h/ were also poorly perceived by the learners. The most misidentified unaspirated dental z /ts/ sound was heard as /s/ 20% of times. The aspirated dental affricate c /ts^h/ sound was misidentified as ch /tʂ^h/ and /s/ 11% each.

3. Experiment 2: Production of Mandarin Consonants

3.1. Participants

The participants were the same 25 beginning level CFL learners who took the perception test in Experiment 1. They provided the production data through a reading task that took place immediately before the perception tests.

3.2. Material and Procedure

The reading list consisted of 20 target words in pinyin embedded in a carrier sentence *wo shuo ___ zi* (我说—字). “I say — word”. Each of the 20 target words was repeated once, yielding two versions of the same target sounds. The participants were given a few minutes to prepare for the reading task. Any questions about the pronunciation of any sounds were answered during the preparation time. The participants were told to read the list at normal speed. The recordings were then made on a MacBook Pro computer using the Praat software. The words containing the eight target consonants

(z/tsa/, c/ts^ha/, j/tc^ja/, q/tc^ha/, x/c^ja/, zh/tʂa/, ch/tʂ^ha/, and sh/ʃa/) were separated from the sentences using waveform editing, normalized for peak volume, and saved as wave form for presentations.

3.3. Assessment

Three phonetically trained native Mandarin speakers, all have taught Mandarin Chinese courses in North America, assessed the participants’ productions in an eight-way forced choice identification task followed by the goodness rating task along a scale of 1 (poor) to 7 (good). The participants’ productions of the eight target sounds were blocked by groups of five speakers, yielding 80 tokens in each session (8 words 5 speakers 2 repetitions). The eight-way forced choice task and the subsequent rating task were created using Praat ExperimentMFC identification test design. Individual identification tasks were carried out on a MacBook Pro computer in a quiet room. The native Mandarin speakers listened to each Mandarin stimulus and identified the initial consonant by clicking on the corresponding label in pinyin on the computer screen. Immediately after the identification of each sound, the listeners rated the fitness of the sound they identified by choosing a number along the scale of 1 (poor) to 7 (good). In the cases of uncertainty, the listener could replay the stimulus up to three times before the choice was made. In addition to the eight target sounds in pinyin, /s/, /t/ and /k/ sounds were also included in the labels for identifications. Based on a screening test by the first author, these three sounds provided additional options for the listeners to choose for the mis-produced target sounds. The listeners all had a trial session to learn the test procedure before the real judgement test began. They each then completed 6 sessions with mandatory breaks in between sessions. The data of one session were excluded from analysis as those five participants were not native English speakers.

3.4. Results

To assess interrater variability, a reliability test was carried out and a high degree of agreement was found among the three raters. The average measures Intraclass Correlation was 0.821 with a 95% confidence interval from 0.788 to 0.849 ($F(399,798) = 5.605, p < 0.001$). Therefore, the mean group production score for each consonant was calculated by taking the average of the three listeners’ identification scores. To further explore the production substitution patterns of each mis-produced consonant, a confusion matrix was created. Table 4 summarizes the mean percentage correct production scores (in bold) and the confusion matrix of the mis-produced eight target consonants. The mean goodness rating scores of each target consonant were also calculated and presented (in italic) by the mean correct identification scores in Table 4.

Table 4. Mean percentage correct production scores (in bold), mean rating scores (in italic), and confusion matrix of mis-produced Mandarin consonants by native English CFL learners (N = 25).

Target	Identified										
	zh /tʂ/	ch /tʂ ^h /	sh /ʃ/	j /tc/	q /tc ^h /	x /c/	z /ts/	c /ts ^h /	/s/	/t/	/k/
zh /tʂ/	25 (4.8)	1	6	37	6	1	19	2	1		
ch /tʂ ^h /	5	47 (5)	1	2	35	1	1	7		2	
sh /ʃ/	4	8	52 (5)	2	7	22	1	2	1		
j /tc/	13	1		68 (5.6)	3	2	12			1	
q /tc ^h /	3	9	1	25	43 (4.7)	7	2	5			4
x /c/	1		7	11	5	64 (5.5)	4	2	6	1	
z /ts/	3			2			88 (5.1)		6		
c /ts ^h /	1	4	1	5	1		27	25 (4.1)	8	7	23

Overall, the percentage correct production scores of the eight target sounds, as identified by the three native listeners, ranged from 25% (zh /tʂ/ and c/ts^h/) to 88% (z /ts/). A One-Way ANOVA on the percentage identification scores revealed a significant effect of consonant (Wilk’s Lambda = 0.288, $F(7, 143) = 50.486, p = 0.000$). The subsequent post hoc Bonferroni tests adjusted for multiple comparisons revealed that a series of pairwise comparisons were significant. The results are presented in Table 5.

Table 5. Pairwise comparisons of differences between the consonants in production (** $p < 0.01$, * $p < 0.05$).

	c /ts ^h /	ch /tʃ ^h /	j /tɕ/	q /tɕ ^h /	sh /ʃ/	x /ç/	z /ts/	zh /tʃ/
c /ts ^h /	-	**	**	*	**	**	**	
ch /tʃ ^h /	**	-	*				**	**
j /tɕ/	**	*	-	**			*	**
q /tɕ ^h /	*		**	-		**	**	*
sh /ʃ/	**				-		**	**
x /ç/	**			**		-	**	**
z /ts/	**	**	**	**	**	**	-	**
zh /tʃ/		**	**	*	**	**	**	-

In addition to the percentage correct identification scores, the assessment of the learners’ production performance also included the rating scores (along a scale of 1 to 7) for each correctly identified consonant. The mean rating scores of each correctly identified sound ranged from 4.1 to 5.5 out of 7. The goodness rating task provided the listeners with a choice among a range of “fitness” of the learner’s production to the native norm of the target sound, even if the intended sound was correctly identified. Therefore, taking into consideration of the rating scores, the “adjusted” production score for each consonant was calculated by multiplying the percentage correct identification score by the rating score. The results of the adjusted production scores for the eight target consonants are summarized in Table 6.

Table 6. Mean adjusted production scores (% production x rating score) of the target consonants produced by English CFL learners (N = 25).

Sounds	Production Score	Rating Score	Adjusted Score
z /ts/	88	5.1	4.5
j /tɕ/	68	5.6	3.8
x /ç/	64	5.5	3.5
sh /ʃ/	52	5	2.6
ch /tʃ ^h /	47	5	2.4
q /tɕ ^h /	43	4.7	2.0
zh /tʃ/	25	4.8	1.2
c /ts ^h /	25	4.1	1.0
Mean	52 (21.5)	5 (.5)	2.6 (1.2)

3.5. Discussion

The two most difficult consonants for the learners to produce were c /ts^h/ and zh/tʃ/, each with a low percentage production score of 25% only, which was significantly different from all the other sounds under investigation. Mandarin c /ts^h/ and zh/tʃ/ also received the lowest adjusted production scores (c /ts^h/ (1.0) and zh/tʃ/ (1.2), when the rating scores were taken into consideration. On the other hand, the adjusted production scores for z /ts/ and j /tɕ/ were the highest among the eight sounds. The mean rating score of the eight target consonants was 5.0, with a range from 4.1 to 5.6. The difference between the best and worst rated sound was 1.5. These rating scores suggest that the listeners did not use the full range of the rating scale, especially at the lower end, once a target sound was correctly identified. While the range of the rating scores was relatively small comparing to the widely different percentage correct identification scores, those three poorly produced sounds with the lowest percentage correct identification scores c /ts^h/ (25%), zh/tʃ/ (25%), and q/tɕ^h/ (35%) also received the rating scores below 5. All the other five sounds had a rating score of 5 and above.

The substitution patterns in production were similar to those in perception for the retroflex sounds. The mis-produced retroflex sounds zh/tʃ/, ch/tʃ^h/, and sh/ʃ/ were overwhelmingly heard as the palatal sounds j/tɕ/, (37%), q/tɕ^h/ (35%), and x/ç/ (22%) by the native Mandarin listeners. However, the mis-produced palatal sounds q/tɕ^h/ and x/ç/ were mostly heard by the native Mandarin listeners

as the unaspirated palatal affricate $j/tʃ/$. For the aspirated dental affricate $c/tsʰ/$, 23% was heard as the $/k/$ sound. This unexpected substitution pattern was more likely caused by the pinyin spelling of “ca” being mistaken as the English orthography. Obviously, these speakers have not learned the $c/tsʰ/$ sound, or, at least have not associated the pinyin c with the Mandarin sound $/tsʰ/$.

4. Perception and Production Comparisons

Visual inspection of Figure 1, which compares the mean percentage correct perception and production scores of the eight consonants, shows the patterns of higher scores in perception than in production for retroflex sounds $zh/tʂ/$, $ch/tʂʰ/$, and $sh/ʂ/$ but vice versa for palatal sounds $j/tʃ/$, $q/tʃʰ/$, and $x/ç/$. The dental affricates were mixed as $c/tsʰ/$, the aspirated dental affricate received the lowest production score of 25% while the unaspirated counterpart $z/ts/$ had the highest production score of 88%.

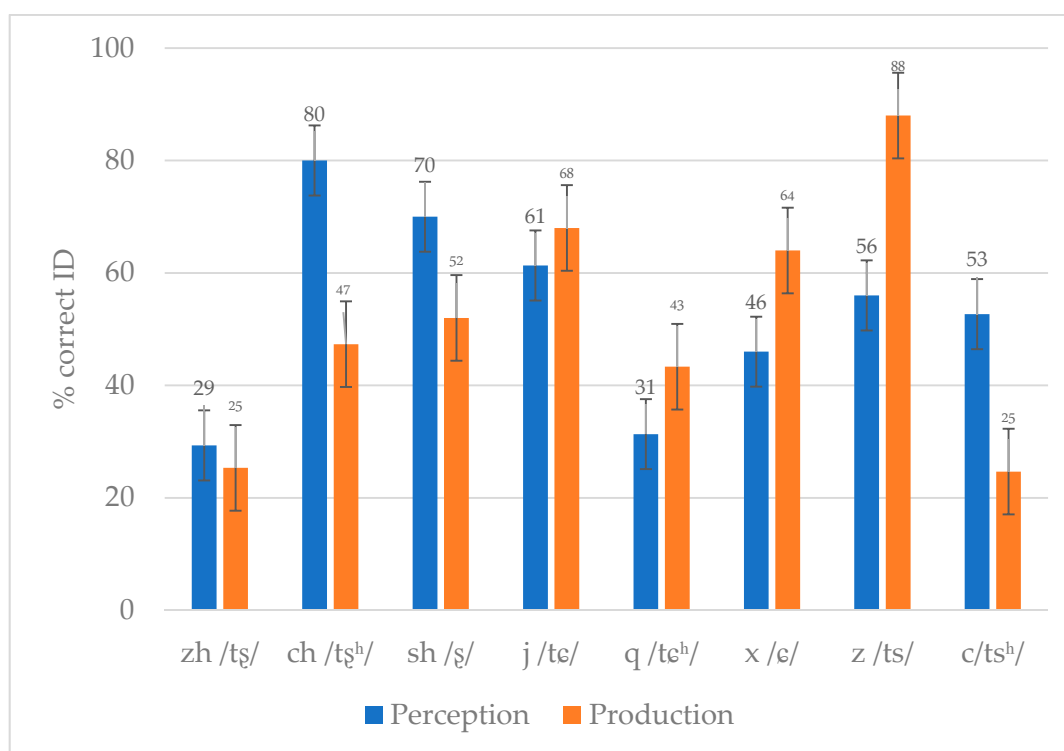


Figure 1. Mean % perception and production scores of Mandarin consonants by English CFL Learners (N = 25).

To investigate the relationship between the perception and production accuracies by the participants, Pearson Coefficients Correlation tests were performed on the percentage correct perception and production scores. The strength of the correlation test results, along with the mean percentage perception and production scores and standard deviations are presented in Table 7.

Results of the Pearson coefficients correlation tests (2-tailed) revealed a moderate size of correlations between the perception and production scores for two of the eight consonants. They were $c/tsʰ/$, $r = 0.619$, $p < 0.01$, and $x/ç/$, $r = 0.508$, $p < 0.01$. No significant correlations between the perception and production scores were found for the remaining six consonants. The Pearson’s r ranged from ($r = 0.028$, $p = 0.896$) for $q/tʃʰ/$ to ($r = 0.077$, $p = 0.715$) for $j/tʃ/$.

Table 7. Pearson coefficients correlation tests between the % correct perception and production scores (** $p < 0.01$).

Sounds.	Perception	Production	r =	p =
zh /tʂ/	29 (29)	25 (34)	0.061	0.773
ch /tʂʰ/	80 (25)	47 (34)	0.039	0.854
sh /ʃ/	70 (25)	52 (37)	0.030	0.888
j /tɕ/	61 (34)	68 (38)	0.077	0.715
q /tɕʰ/	31 (21)	43 (35)	0.028	0.896
x /ç/	46 (29)	64 (37)	0.508 **	0.010
z /ts/	56 (28)	88 (24)	0.162	0.439
c /tsʰ/	53 (36)	25 (30)	0.619 **	0.001

5. General Discussion and Conclusions

To answer research question 1 which asked which Mandarin consonants pose difficulties for native English CFL learners, results of Experiment 1 showed that the learners had different degrees of difficulties in identification of the eight Mandarin consonants under investigation. The most difficult sounds were zh /tʂ/ (29%), q /tɕʰ/ (31%), and x /ç/ (46%). The findings were consistent with an earlier study by Wang and Chen (2019) in which zh /tʂ/, q /tɕʰ/, and x /ç/ were also among the four most difficult categories identified by the low level CFL learners. Similarly, Yang and Yu (2019) also found that among the six Mandarin affricates they investigated, zh /tʂ/ and q /tɕʰ/ were more difficult than their counterparts ch /tʂʰ/ and j /tɕ/ for the native English CFL learners.

The confusion matrix of misidentified sounds showed the English CFL learners substituted the retroflex and palatal sounds with each other mostly and such confusion patterns suggest that the learners have not established separate categories for these sounds. The English CFL learners' problems with the retroflex and palatal sounds, also reported in Hao (2012) and Yang and Yu (2019) studies, are closely related to the phonetic differences between their L1 and L2 sound systems. Wang and Chen (2019) found the native English listeners mapped both Mandarin retroflex affricates zh/tʂ/, ch /tʂʰ/, and the aspirated palatal affricate q/tɕʰ/ onto the English /tʃ/ sound, though the degree of "fitness" was different, as ch /tʂʰ/ was identified as a much better fit to /tʃ/ (4.4) than zh/tʂ/ (3.3), and q/tɕʰ/ (2.3), indicated by their "fit indexes": (% identification × goodness rating score). The three-to-one perceptual assimilation pattern, to a large extent, underlies the native English CFL learners' perception problems with zh /tʂ/, ch /tʂʰ/, and q /tɕʰ/. The better fitting category ch /tʂʰ/ was identified with an accuracy score of 80%, as compared with 31% for q/tɕʰ/ and 29% for zh/tʂ/ in the current study. The findings support the Category Goodness (CG) type of assimilation of the Perceptual Assimilation Model (PAM), which states that two sounds are assimilated to a single native category resulting in a better fit for one than the other (Best et al. 2001). The current findings suggest that the two-to-one CG type of assimilation can be expanded to three-to-one assimilation. More such three-to-one, and two-to-one, as well as one-to-two mappings of Mandarin consonants onto English categories found in the Wang and Chen (2019) study are presented in Figure 2. These cross-linguistic assimilation patterns shed light on the difficulties native English CFL learners demonstrated in their perception and production of the eight target consonants in the current findings. (See the original study for detailed analysis of the assimilation patterns).

L2 to L1 assimilation patterns can also explain the English CFL learners' difficulties with Mandarin palatal sound x/ç/. Both x/ç/ and sh/ʃ/ were mapped onto English /ʃ/ but sh/ʃ/ was a better fit than x/ç/, a CG type of assimilation. Hao (2012) reported the same CG assimilation pattern for x/ç/ and sh/ʃ/ and similar learning results on x/ç/ in her study. Mandarin x/ç/ was difficult for the English learners also because it was assimilated to both English /z/ and /ʃ/, a one-to-two "split" match to the native categories, a "revised" Single Category (SC) type of assimilation (see Figure 2).

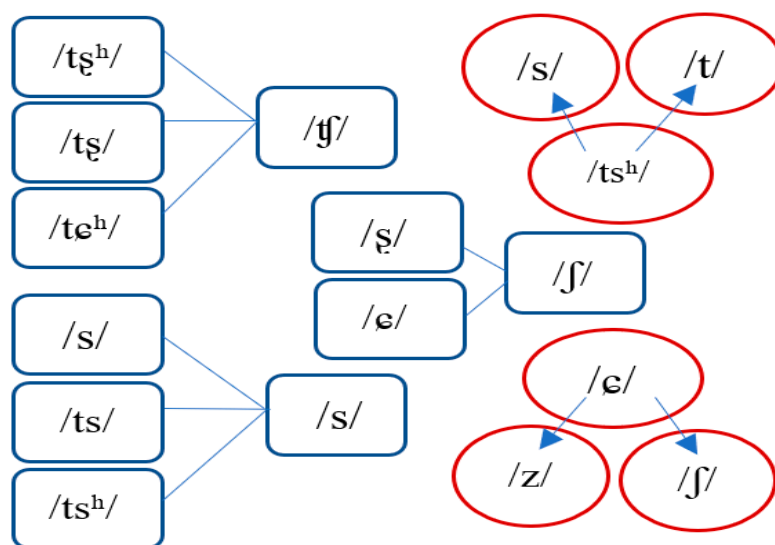


Figure 2. Mandarin to English sound mapping patterns by English listeners; 3 to 1 and 2 to 1 mappings are in blue squares and 1 to 2 mappings are in red circles.

The Mandarin dental affricates *z* /ts/ and *c* /tsʰ/ were also poorly perceived by the learners in the current study. As seen in Figure 2, the aspirated affricate /tsʰ/ was assimilated to both English /s/ and /t/, also a “revised” Single Category (SC) type of assimilation. The unaspirated dental affricate *z* /ts/, together with *s* /s/ and *c* /tsʰ/, were mapped onto English /s/, causing difficulties for the poor fitting categories *z* /ts/ and *c* /tsʰ/. While both *z* /ts/ and *c* /tsʰ/ exist in the English word finals “reads” and “boots”, these novice learners did not appear to have made the associations in identifying the target sounds. One explanation may be that Mandarin *z* /ts/ and *c* /tsʰ/ are stand-alone phonemes and are more prominent at word initial positions than the morphological word endings of /dz/ and /ts/ in English.

Overall, to answer research question 2, the current data suggest that the perceived phonetic differences and distances between Mandarin and English consonants predicted the learners’ perceptual difficulties with the L2 Mandarin consonants. The perception data also support the PAM model.

Flege’s Speech Learning Model (SLM) may also provide explanations for the current findings. The learners’ phonetic spaces for L1 and L2 consonants need to be reorganized to establish new phonetic categories for the Mandarin retroflex, palatal and dental sounds. For example, learners need to distinguish the differences between *c* /tsʰ/, *x* /ç/, *z* /ts/, *q* /tʃʰ/ and others in order to establish these categories. On the other hand, “equivalence classification” of the SLM may be at work for *ch* /tʂʰ/ to be identified as English /tʃ/. While *ch* /tʂʰ/ (80%) was the best identified category among the eight target sounds by the learners, its production score (47%) was much lower. Therefore, even if some of the L2 categories seemed to have been established by the majority of the listeners, “equivalence classifications” may have prevented them from forming the native-like perception category.

The results of Experiment 2 showed the percentage correct production scores of the eight target sounds ranged from 25% (*zh* /tʂ/ and *c* /tsʰ/) to 88% (*z* /ts/). Pairwise comparisons data shown in Table 4 indicated the native English CFL learners had the most production difficulties with the Mandarin *c* /tsʰ/ and *zh* /tʂ/, followed by *q* /tʃʰ/ sounds. The pattern of substitutions in production was similar to that of perception for the retroflex sounds *zh* /tʂ/, *ch* /tʂʰ/, and *sh* /ʃ/, which were substituted with palatals *j* /tʃ/, *q* /tʃʰ/, and *x* /ç/. However, the mis-produced palatal sounds *q* /tʃʰ/, *x* /ç/ were not confused with the retroflex sounds but were mostly heard by the native Mandarin listeners as the unaspirated palatal affricate *j* /tʃ/. These substitution patterns suggest that the retroflex sounds were more difficult for the English CFL learners to produce.

Comparing the results of the two experiments, the learners had the tendency of better performance on the retroflex sounds *zh* /tʂ/, *ch* /tʂʰ/, and *sh* /ʃ/ in perception than in production but vice versa on

palatal sounds $j/tɕ/$, $q/tɕʰ/$, and $x/ç/$. The results on dental sounds $z/ts/$, and $c/tsʰ/$ were mixed across the two domains. Mandarin retroflex and palatal fricatives and affricates, though both lack counterparts in English, pose different problems to the English CFL learners in perception and production. The results of the correlation tests comparing the perception and production scores showed only two of the eight target consonants, $x/ç/$ and $c/tsʰ/$ were moderately correlated. The lack of correlations in the learners' perception and production scores for the majority of the sounds under investigation suggest the relationship between L2 speech perception and production is not straightforward.

One possible explanation for such misalignment in L2 speech perception and production might be that perception does not always lead production. For example, different mechanisms or strategies may be involved in perception and production of the retroflex sound $ch/tʂʰ/$ by the beginning level English CFL learners. The learners' better perception of $ch/tʂʰ/$ (80%) may be explained by their closest match of the L2 Mandarin $ch/tʂʰ/$ to their L1 English $/tʃ/$ sound. In perception identification tasks, anything that is close enough to the nearest English $/tʃ/$ can be labeled as $ch/tʂʰ/$. However, the same strategy would not work for the production, if the learners have not established native-like retroflex $ch/tʂʰ/$ category. The key phonetic gestures in producing the correct retroflex affricates in the Mandarin $ch/tʂʰ/$ sound cannot be effectively replaced by the gestures of English $/tʃ/$. The intended $ch/tʂʰ/$ would not be heard as the target $ch/tʂʰ/$ sound but as the $q/tɕʰ/$ sound by the native Mandarin listeners. The same patterns seem to hold true for the other retroflex sounds $zh/tʂ/$ and $sh/ʂ/$ that were identified by the native Mandarin listeners as the palatal sounds $j/tɕ/$ and $q/tɕʰ/$. These substitution patterns suggest the cues for the retroflex sounds were absent in these non-native productions. Therefore, it is very likely the misalignment between perception and production is partly due to the different mechanisms the learners attended to in perception and production of the target consonants. Acoustic analyses of the learners' productions of these sounds, along with perceptual test using synthesized stimuli manipulating the key acoustic cues differentiating the target categories are needed to draw a firm conclusion in future studies.

The current data also show that partial alignment between perception and production of Mandarin consonants does exist. The listeners' perception and production scores of both $x/ç/$ and $c/tsʰ/$ were moderately but significantly correlated, indicating the link between perception and production. Past studies have come to the same conclusions of such partial alignment in perception and production of L2 sounds (Flege 1999; Rochet 1995).

L2 phonetic training studies have also examined the relationship between perception and production when assessing the outcomes of the training. There is evidence that perceptual training only led to improvement in both perception and production of L2 consonants (Bradlow et al. 1997) and lexical tones (Wang 2008, 2012, 2013; Wang et al. 2003), and production gains are larger on obstruents than on sonorants and vowels (Sakai and Moorman 2018). There is also evidence that trainees' perceptual learning did not lead to better productions on L2 vowel contrasts (Wang 2002). These findings suggest that the relationship between perception and production of L2 speech sounds can be further complicated by the different sound classes.

In conclusion, the current data showed partial alignment but more discrepancies on native English CFL learners' perception and production of the eight Mandarin consonants. Different phonetic mechanisms and strategies may be involved in the L2 speech sound perception and production. Perception may not always lead production. Future studies need to carry out more detailed acoustic analysis of the native and non-native productions of the target consonants to investigate specific problems that learners have in perception and production of Mandarin consonants.

Finally, one limitation of the current study was the exclusion of the $s/s/$ and $r/ʐa/$ sounds in the analyses. It would have been better to include at least the $s/s/$ sound to have a nice set of dental fricative and affricates, making it parallel to the retroflex and palatal sets.

Author Contributions: Conceptualization, X.W. and J.C.; methodology, X.W. and J.C.; data collection and analysis, X.W. and J.C.; writing—original draft preparation, X.W.; writing—review and editing, X.W. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research, Scholarly and Creative Activities (RSCA) award to the first author (2018–2019) by California State University, Fresno.

Acknowledgments: For data collection, we are grateful to Li Mann, Xinping Yu, and Pei Chen Chang for their support and their permission for us to recruit their students. We also thank all the participants for their participation.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Best, Catherine. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Edited by Judith C. Goodman and Howard C. Nusbaum. Cambridge: MIT Press, pp. 167–244.
- Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America* 109: 775–94. [[CrossRef](#)]
- Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299–310. [[CrossRef](#)]
- Chuang, Yu-Ying, Ching-Chu Sun, Janice Fon, and R. Harald Baayen. 2019. Geographical Variation of the Merging between Dental and Retroflex Sibilants in Taiwan Mandarin. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Canberra: Australasian Speech Science and Technology Association Inc., pp. 472–76.
- Evans, Bronwen G., and Wafaa Alshangiti. 2018. The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics* 68: 15–31. [[CrossRef](#)]
- Flege, James. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–77.
- Flege, James. 1999. The relation between L2 production and perception. In *Proceedings of the XIVth International Congress of Phonetic Sciences*. Edited by John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey. College Park: American Institute of Physics, pp. 1273–76.
- Flege, James. 2007. Language contacts in bilingualism: Phonetic systems interactions. *Laboratory Phonology* 9: 353–80.
- Flege, James E., and Ratre Wayland. 2019. The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies* 2: 1–44. [[CrossRef](#)]
- Guion, Susan G., James E. Flege, Reiko Akahane-Yamada, and Jessica C. Pruitt. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America* 107: 2711–24. [[CrossRef](#)] [[PubMed](#)]
- Hao, Yen-Chen. 2012. The Effect of L2 Experience on Second Language Acquisition of Mandarin Consonants, Vowels, and Tones. Ph.D. dissertation, Indiana University, Bloomington, IN, USA.
- Lai, Yi-hsiu. 2009. Asymmetry in Mandarin affricate perception by learners of Mandarin Chinese. *Language and Cognitive Processes* 24: 1265–85. [[CrossRef](#)]
- Lin, Hua. 2005. Understanding problems in learning Mandarin consonants by monolingual speakers of English. *Journal of Canadian Teachers of Chinese as a Second Language* 1: 1–18.
- Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurement. *Word* 20: 384–422. [[CrossRef](#)]
- Lisker, Leigh, and Arthur S. Abramson. 1970. The voicing dimension: Some experiments on comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia, vol. 563, pp. 563–77.
- Liu, Jiang, and Allard Jongman. 2012. American Chinese learners' acquisition of L2 Chinese affricates/ts/and/tsh/. In *Proceedings of Meetings on Acoustics 164ASA*. Melville: Acoustical Society of America, vol. 18, p. 060005.
- Munro, Murray, and Tracey M. Derwing. 2008. Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning* 58: 479–502. [[CrossRef](#)]

- Munro, Murray, Tracey M. Derwing, and Ron I. Thomson. 2015. Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching* 53: 39–60. [CrossRef]
- Polka, Linda, and Ocke-Schwen Bohn. 1996. A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America* 100: 577–92. [CrossRef] [PubMed]
- Rochet, Bernard. 1995. Perception and production of second-language speech sounds by adults. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 379–410.
- Sakai, Mari, and Colleen Moorman. 2018. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics* 39: 187–224. [CrossRef]
- Strange, Winifred. 1995. Cross-language studies of speech perception: A historical review. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 3–45.
- Wang, Xinchun. 1997. The Acquisition of English Vowels by Mandarin ESL Learners: A Study of Production and Perception. Master's thesis, Simon Fraser University, Burnaby, BC, Canada.
- Wang, Xinchun. 2002. Training Mandarin and Cantonese Speakers to Identify English Vowel Contrasts: Long-Term Retention and Effects on Production. Doctoral dissertation, Simon Fraser University, Burnaby, BC, Canada.
- Wang, Xinchun. 2006. Perception of L2 tones: L1 lexical tone experience may not help. Paper presented at Third International Conference on Speech Prosody, Dresden, Germany, May 2–5; pp. 85–88.
- Wang, Xinchun. 2008. Training for learning Mandarin tones. In *Handbook of Research on Computer-Enhanced Language Acquisition and Learning*. Edited by Felicia Zhang and Beth Barber. Hershey: IGI Global, pp. 259–74.
- Wang, Xinchun. 2012. Auditory and Visual Training on Mandarin Tones: A Pilot Study on Phrases and Sentences. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)* 2: 16–29. [CrossRef]
- Wang, Xinchun. 2013. Perception of Mandarin Tones: The Effect of L1 Background and Training. *The Modern Language Journal* 97: 144–160. [CrossRef]
- Wang, Xinchun, and Jidong Chen. 2019. English Speakers' Perception of Mandarin Consonants: The Effect of Phonetic Distances and L2 Experience. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Canberra: Australasian Speech Science and Technology Association Inc., pp. 250–54.
- Wang, Xinchun, and Murray Munro. 2004. Computer-based training for learning English vowel contrasts. *System* 32: 539–52. [CrossRef]
- Wang, Yue, Allard Jongman, and Joan A. Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *The Journal of the Acoustical Society of America* 113: 1033–43. [CrossRef] [PubMed]
- Werker, Janet. 1994. Cross-language speech perception: Developmental change does not involve loss. In *The Development of Speech Perception: The Transition from Speech sounds to Spoken Words*. Edited by Judith C. Goodman and Howard C. Nusbaum. Cambridge: The MIT Press, pp. 93–120.
- Werker, Janet, and Linda Polka. 1993. Developmental changes in speech perception: New challenges and new directions. *Journal of Phonetics* 21: 83–101. [CrossRef]
- Yang, Chunsheng, and Alan C. L. Yu. 2019. The acquisition of Mandarin affricates by American L2 learners. *Taiwan Journal of Linguistics* 17: 91–122.
- Zhu, Liyi. 2012. Retroflex and Non-Retroflex Merger in Shanghai Accented Mandarin. Master's thesis, University of Washington, Seattle, WA, USA.

