English Speakers' Perception of Mandarin Consonants: The Effect of Phonetic Distances and L2 Experience

Xinchun Wang, Jidong Chen

California State University, Fresno

ABSTRACT

Two experiments were conducted to investigate the influence of the perceived phonetic distances between Mandarin and English consonants on L2 consonant identification by English learners of Chinese as a foreign language (CFL). Experiment 1 elicited English listeners' identification of ten target Mandarin consonants as English categories and their goodness fitting ratings of each mapping. Fit indexes, measuring the degree of mapping between Mandarin and English consonants, ranged from poor to fair, and good. Experiment 2 elicited English CFL learners' (elementary and intermediate levels) identification of Mandarin consonants in a forced choice task. Overall, the perceived phonetic distances between Mandarin and English consonants predicted the learners' correct identification of the L2 consonants. Mandarin consonants with high fit indexes to English consonants are better identified than those with lower fit indexes. More L2 experience also leads to better perceptual learning. The findings are discussed in terms of current L2 speech learning models.
Keywords: Mandarin consonants, cross-language sound classification, L2 speech perception

1. INTRODUCTION

Research has shown that phonetic differences and distances between L1 and L2 speech sounds is an important factor that contributes to the degree of success in perception of L2 speech sounds [1, 2]. The two most influential L2 speech perception and learning models, Best's Perceptual Assimilation Model (PAM) [3,4] and Flege's Speech Learning Model (SLM) [5,6] both assume that learners' perceptual assimilation or dissimilation patterns of L2 sounds to L1 categories is systematically related to their native phonetic system. According the PAM model, several pairwise assimilation types are possible when the two non-native phones are mapped on the L2 system. The L2 phones may be assimilated 1) to two different L1 phones, the Two Category (TC) type, or 2) to a single L1 category, the Single Category type (SC) equally poor or well, or, 3) to a single native category but one is a better fit than the other, the Category Goodness type (CG). PAM also predicts the "gradients" of difficulties in discriminations of L2 sounds from the most to the least difficulties: SC > CG >TC) [4]. On the other hand, SLM posits that speakers' L1 and L2 sound systems interact and exist in a common phonological space. Whether new L2 phonetic categories are established or not depends on the perceived dissimilarities of an L2 sound from the closest L1 or L2 sounds. Learners' ability to establish such new phonetic categories increases with increased L2 experience. Equivalence Classification actually blocks the formation of new L2 categories [5, 6].

With regard to the methodologies of assessing phonetic distances between L1 and L2 speech sounds, the commonly used method of phoneme inventory comparisons is not sufficient as the IPA symbols do not provide the detailed phonetic properties of sounds across languages. Predictions based on phonetic/ acoustic properties of phones across languages may not be sufficient either as such measurement may not capture the most crucial phonetic cues of category formation. Cross-language speech perception, that is, having the listeners identify the target L2 sounds as their L1 categories, adopted in recent L2 speech research, is a more reliable method [7]. The current study uses such cross-linguistic perceptual mapping method to assess the phonetic distances between Mandarin and English consonants.

Table 1: Mandarin Consonants in IPA

	Labial	Dental-Alveolar	Retroflex	Alveolo-palatal	Velar
Stop	p pʰ	t tʰ			k kʰ
Affricate		ts tsʰ	tʂ tʂʰ	tɕ tɕʰ	
Fricative	f	s	ʂ ʐ	ɕ	x
Nasal	m	n			ŋ
Liquid		l			

The target sounds under investigation (bold faced in Table 1) are ten Mandarin consonants in *pinyin* and IPA symbols: z /ts/, c /tsʰ/, s /s/, j /tɕ/, q /tɕʰ/, x /ɕ/, zh /tʂ/, ch /tʂʰ/, sh /ʂ/, r /ʐ/. They form the fricative/affricate groups at dental, retroflex, and Alveolo-palatal places reported to be difficult for English CFL learners in both perception and production as most of these sounds do not have corresponding counterparts in English [8]. There is a paucity of literature on L2 perception of Mandarin consonants and, to the knowledge of the researchers, no study on direct mapping of Mandarin consonants

onto English categories through cross-linguistic identification test.

Learners' L2 learning experience also plays an important role in the success of acquisition of L2 sounds. Numerous past studies have indicated that experienced learners performed significantly better in mastering the perception and production of L2 sounds [2]. The current study also investigates the effect of L2 experience by comparing beginning and intermediate level CFL learners' perceptual accuracy of Mandarin consonants. The research questions are: 1) How do the perceived phonetic distances between L1 and L2 sounds influence English CFL learners' perceptual identification of Mandarin consonants? 2) Does increased L2 experience make a difference in perceptual learning of L2 Mandarin consonants?

2. EXPT. 1: CROSS- LINGUISTIC PERCEPTUAL CLASSIFICATION

2.1. Participants

The participants were 16 (6 male, 10 female) native English-speaking undergraduate students at a U.S. university. Six of the participants reported speaking another language as their first language along with English (3 Spanish, 2 Punjabi, and 1 Hmong). Two of them were born in a foreign country but moved to the U.S. at a very young age. The mean age of the group is 19.5 years (range: 18-24). The participants had some basic training in linguistics and all were enrolled in an introduction to Linguistics course at the time of the study. All participants have taken foreign language courses, mostly Spanish at high school but none have studied Mandarin Chinese.

2.2. Material

Ten Mandarin Chinese consonants in CV position produced by a male native speaker were used as stimuli for the cross-linguistic identification task. The target words were produced in a carrier sentence *wo shuo ___ zi* (我说 ---字). "I say --- word". The recordings were made on a MacPro computer using Praat software. The target words were separated from the sentences using waveform editing, normalized for peak volume, and saved as wave form for presentations.

2.3. Procedure

The listening tasks were performed in a classroom equipped with an internal speaker system. The participants were given detailed instructions about the identification and rating tasks. The instruction was followed by a practice session using 11 Mandarin

consonants (b p m f d t n l g k h) in CV syllables (V is the low vowel /a/) (not included in the analysis) to familiarize the listeners with the identification and rating tasks. The 10 test stimuli in the same C+/a/ syllables (z /ts/, c /tsʰ/, s /s/, j /tɕ/, q /tɕʰ/, x /ɕ/, zh /tʂ/, ch /tʂʰ/, sh /ʂ/, r /ʐ/) were randomized and played back three times with an inter stimulus interval (ISI) of 7 seconds for identification and goodness rating tasks. The participants listened to each Mandarin stimulus once and identified it as one of the ten English sounds by circling the corresponding item on the answer sheet. Immediately after the identification of the stimulus sound, the listeners rated the fitness of the sound they identified by circling a number along the scale of 1 (poor) to 7 (good). The 10 English sounds chosen for the identifications are cha sha sa ra ja za ta da θa and ða ('tha' is avoided because it could not distinguish the voice contrasts). The choices were made based on the results of a pilot test on two trained native English-speaking phoneticians.

2.4. Results

Table 2 presents the mean percentages of identifications of the 10 Mandarin consonants as English sounds along with the rating scores. The number in boldface is the "modal classification", indicating the highest frequency of identifications of each Mandarin consonant as the English category. To take into account both the identification and the rating scores for the measurement of the perceived phonetic distances between English and Mandarin sounds, the *fit index* was calculated for each cross linguistically classified category that received more than 25% of identification score (see Table 2).

Table 2: Mean % ID, rating scores, and fit indexes of Mandarin to English sound mapping by English listeners.

Mandarin Sounds	English ID	% ID	Rating	Fit Idx	Match
r /ʐ/	/ɹ/	**100**	6.3	6.3	Good
sh /ʂ/	/ʃ/	**100**	5.8	5.8	Good
ch /tʂʰ/	/tʃ/	**90**	4.9	4.4	Fair
s /s/	/s/	**77**	5.6	4.3	Fair
z /ts/	/s/	**79**	4.9	3.9	Fair
j /tɕ/	/dʒ/	**81**	4.4	3.6	Poor
zh /tʂ/	/tʃ/	**67**	4.9	3.3	Poor
q /tɕʰ/	/tʃ/	**67**	3.5	2.3	Poor
c /tsʰ/	/s/	**46**	3.8	1.7	Poor
c /tsʰ/	/t/	38	4.5	1.7	Poor
x /ɕ/	/ʃ/	**46**	2.6	1.2	Poor
x /ɕ/	/z/	27	3.7	1.0	Poor

The fit index was derived from multiplying the percentage of (the proportion of) identifications and goodness ratings. As seen in Table 2, there is a range of phonetic distances between the L1 and L2 sounds based on the fit indexes (6.3-1.0). As a working hypothesis, the 10 categories were divided into "good", "fair", and "poor" subgroups based on the *mean* fit indexes (3.7, s.d.=1.7) of the *modal classification* of each sound. The "poor" matching categories were x /ɕ/, c /tsʰ/, q /tɕʰ/, zh /tʂ/, and j /tɕ/ whose fit indexes were *below* the mean. The "fair" fitting categories were ch /tʂʰ/, s /s/, and z /ts/ whose fit indexes were at or above the mean. The "good" matching sounds were r /ʐ/, sh /ʂ/ whose fit indexes were 1s.d. above the mean.

In terms of perceptual assimilation patterns, although eight of the 10 Mandarin consonants had the modal classifications of 67% -100% of instances, they were not all simple one on one mapping categories (see Figure 1). Three Mandarin categories were all heard as the English /ʧ/ but with different fit indexes: q /tɕʰ/ (2.3), zh /tʂ/ (3.3) and ch /tʂʰ/ (4.4), indicating /tʂʰ/ was a "fair" match and both zh /tʂ/ and q /tɕʰ/ were "poor" matches for English /ʧ/. Similarly, three Mandarin sounds, s /s/ (4.3), z /ts/ (3.9), and c /tsʰ/ (1.7) were all mapped onto one English category /s/, forming "fair" matches for the former two and "poor" match for the later. Mandarin sh /ʂ/ and x /ɕ/ were both classified as English /ʃ/, with the former a "good" match (5.8) and latter a "poor" match (1.2) category. In contrast to the above 3 to 1 and 2 to 1 mapping patterns, Mandarin c /tsʰ/ and x /ɕ/, appeared to have the 1 to 2 mapping pattern as each one was matched to two English sounds. Figure 1 presents these split matches.

Figure 1: 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.
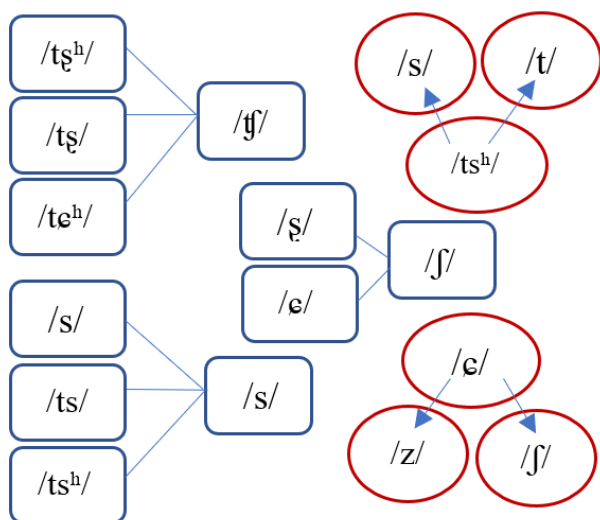
3. EXPT. 2: ENGLISH CFL LEARNERS' PERCEPTION OF MANDARIN CONSONANTS

3.1. Participants

A total of 47 English-speaking CFL learners at a US university at three different levels participated as listeners. The beginning level group consisted of 32 real beginners (18 male, 14 female, mean age = 18.9) enrolled in a first semester Chinese class. At the point of data collection, they were about 3 months into the 16-week semester. Eight participants (3 male, 5 female, mean age = 21.9) enrolled in a third semester Chinese class formed the Early Intermediate Group and the remaining seven learners (3 male, 4 female, mean age=19.3) enrolled in a 5ᵗʰ semester Chinese class formed the Late Intermediate Group. Although all reported speaking English as their first language, some speakers reported speaking another language along with English as their first languages (Hmong, Spanish, Telugu, and Lao).

3.2. Material

The stimuli for the perceptual identification test were the same sounds used in Experiment 1.

3.3. Procedure

The test procedures were the same as Experiment 1 except that the CFL learners identified each Mandarin consonant stimulus by circling the *corresponding item in pinyin* on the answer sheet. The ISI was reduced to 6 seconds as there was no rating task following the identification task.

3.4. Results

The listeners' percentage correct identifications were submitted to a multivariate analysis with Group (Beginning, Early and Late Intermediate) as between group factor and Consonant (10) as multivariate factor. The effect of group was significant (F= 10.229, p=.000.) However, post hoc Tukey HSD tests revealed the differences were significant between the Beginning and the Early Intermediate groups, and between the Beginning and the Late Intermediate groups but not significant between the Early and Late Intermediate Groups. As a result, the latter two groups were combined to form the Intermediate Group for further analysis.

Figure 2 presents the percentage of correct identifications of the Mandarin consonants by the Beginning and Intermediate groups. A One-way ANOVA revealed the mean difference between the two groups was significant F= 17.146, p =.000. A

series of one-way ANOVAs established significant differences between the two groups on zh /tʂ/, F= 4.16, p=.047, q /tɕʰ/, F= 7.463, p =.009, c /tsʰ/, F=15.67, p = .000, x /ɕ/, F= 8.179, p=.006, and on /s/, F=4.889, p = .032. The differences between the other sounds were not significant.

Figure 2: The % correct identifications of Mandarin consonants by the beginning and intermediate groups
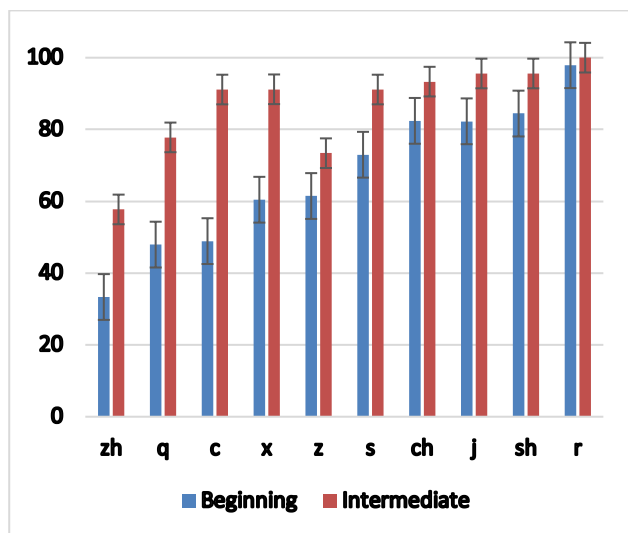
4. DISCUSSION

Experiment 1 shows that the 10 Mandarin consonants were mapped onto English sounds with a range of fit index scores of 1.0 (poorest) to 6.4 (best) by native English listeners. The good matches were r /ʐ/ and sh /ʂ/ to English /ɹ/ and /ʃ/. It is worth noting that sh /ʂ/, a retroflex sound, was 100% identified as English /ʃ/, a surprising finding. The shared feature of friction noise might be the main cue that led to the match of /ʂ/ to the nearby English /ʃ/. Detailed analyses of acoustic properties and perceptual tests that manipulate these acoustic cues is needed to further explore the key acoustic weights that cue the cross-linguistic mapping of /ʂ/ to /ʃ/ and the L2 to L1 matches of all other nine sounds under investigation.

Among the three "fair" matching sounds ch /tʂʰ/, s /s/, and z /ts/, Mandarin /s/, which has an English counterpart, was identified as English /s/ 77% of instances. The problem with /s/ might be related to the confusions caused by the two Mandarin affricates, z /ts/ and c /tsʰ/, the competing matches to English /s/ (see Figure 1). The poor fitting sounds are the Mandarin x /ɕ/, c /tsʰ/, q /tɕʰ/, zh /tʂ/, and j /tɕ/.

How do the perceived phonetic distances between L1 and L2 sounds influence English CFL learners' perceptual identification of Mandarin consonants? Experiment 2 data showed that zh /tʂ/, q /tɕʰ/, c /tsʰ/, and x /ɕ/ received the lowest % identification scores

among the 10 target sounds, especially for the beginning level learners. These four sounds were also the "poor", (also the worst) fitting categories in cross-linguistic mapping test in Experiment 1. On the other hand, the two best matching categories, r /ʐ/, and sh /ʂ/, received the highest % identifications for both groups, followed by the "fair" match sounds ch /tʂʰ/, s /s/, and z /ts/. Therefore, the findings suggest that the perceived phonetic distances between L1 and L2 consonants predicted the CFL learners' L2 Mandarin consonant perception problems, especially for beginning level learners. Revisiting research question 2, which asked whether increased L2 experience makes a difference in perceptual learning of L2 Mandarin consonants, the results showed the intermediate level group performed significantly better than the beginning level group on 5 of the 10 Mandarin consonants, indicating increased L2 learning experience helped the learners in their perceptual identifications of the "poor" fit categories.

With regard to L2 perception theories, the Category Goodness type of the PAM model may explain the sh /ʂ/ and x /ɕ/ to English /ʃ/ match. In fact, the data suggest the CG type may be expanded to include the 3 to 1 matches (see Figure 1): ch /tʂʰ/, zh /tʂ/, and q /tɕʰ/ to English /tʃ/, and s /s/, z /ts/, and c /tsʰ/ to English /s/. In all these CG cases, the better fit sounds (among the 3 or 2 to 1 matches) ch /tʂʰ/, s /s/, and sh /ʂ/ received higher % identification scores than the poorer match categories by the CFL learners. The findings support the predictions of the CG type of the PAM. However, what might be difficult to explain with the PAM model is c /tsʰ/, which is mapped onto two English categories /s/ and /t/ with the same fit index score of 1.7. Similarly, x /ɕ/ categorized as English /ʃ/ (1.2) and /z/ (1.0) also had the 1 to 2 split categorizations (see Figure 1). These may be the "reversed" Single Category type of assimilation in which the target sound is classified as two different L1 sounds. Both sounds proved to be difficult in perception by the CFL learners.

Flege's SLM, a more dynamic model, may also explain some of the current findings. The learners' phonetic spaces for L1 and L2 consonants need to be reorganized to establish new phonetic categories for the poor match Mandarin consonants, especially those 3 to 1, and 2 to 1, as wells as the split mapping sounds discussed in the above. For example, learners need to establish separate categories for c /tsʰ/, x /ɕ/, z /ts/, q /tɕʰ/ and others. On the other hand, "equivalence classification" of the SLM may be at work for the Mandarin r /ʐ/ classified as English /ɹ/ (6.3), and sh /ʂ/ as /ʃ/ (5.8). While these sounds were the best identified categories by the learners, production data are needed to assess the learners' success in producing these categories. The current

data also suggest that L2 learning experience is a predictor for learning, which also supports the SLM.

5. CONCLUSIONS

The cross-linguistic perceptual identification and goodness rating tasks by native English listeners established different assimilation patterns of L2 Mandarin categories to the L1 English consonant system with poor, to fair, and good matches. The data suggest that phonetic distances between Mandarin and English consonants is a predictor for learners' success in perception of Mandarin consonants. Mandarin sounds with higher fit indexes to English categories are better identified by learners than those with lower fit indexes. L2 experience is also an important factor for perceptual learning of Mandarin consonants. Both the PAM and SLM models partially explain the CFL learner's perceptual difficulties.

The current study provided new data to the field of L2 speech perception and filled the gap of cross-linguistic perceptual classification of L2 Mandarin sounds in terms of L1 English sounds. One limitation of the current study is the lack of production data. Future studies need to examine learners' production of Mandarin consonants to gain better understanding of the relationship between the perception and production in L2 consonant acquisition.

7. REFERENCES

[1] Wang, X., Munro, M. 2004. Computer-based training for learning English vowel contrasts. *System 4*, 539-552.

[2] Strange, W. 1995. Cross-language studies of speech perception: A historical review. In: Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 3-45.

[3] Best, C. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, J., Nusbaum, H. (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge: MIT Press, 167-244.

[4] Best, C. T., McRoberts, G. W., & Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoust. Soc. Am.*, *109*, 775-94

[5] Flege, J. 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 233-277.

[6] Flege, J. 2007. Language contacts in bilingualism: Phonetic systems interactions. In: Cole, J., & Haulde, J. (Eds), *Laboratory Phonology 9*. Berlin: Mouton de Gruyder, 353-380.

[7] Guion, S., Flege, J., Akahane-Yamada, R., Pruitt, J. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *J. Acoust. Soc. Am.* 107, 2711–2724.

[8] Lin, H. 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.