

Vowel Effects on L2 Perception of English Consonants by Advanced Learners of English

Yizhou Lan

Shenzhen University

3688 Nanhai Avenue, Nanshan District,

Shenzhen

yzlan@szu.edu.cn

Abstract

A perceptual identification experiment has been designed to study the effects of vowel context in Mandarin speakers' perceptual identification of L2 English fricatives and affricates. The identification task elicited the preferred Mandarin equivalent and a fitness rate of each of about 70 stimuli words with these English consonants from 67 Mandarin-speaking advanced learners of English (advanced EFL learners). The degree of mapping between Mandarin and English consonants, ranging from poor to fair, and good, were compared against predictions by the Perceptual Assimilation Model, a theoretic model that predicts learning outcomes by phonetic distances. Overall, the perceived phonetic distances between Mandarin and English consonants predicted the learners' correct identification of the L2 consonants except for a few number of individual sounds, which showed enormous vowel variation. The variation of mapping patterns across vowel context data and individual sound results suggests that factors other than phonetic similarity, such as articulatory gestures should be accounted for in the prediction of L2 learning outcomes. The findings are discussed along the lines of gestural economy, present L2 speech learning models and pedagogical applications in EFL classrooms.

1 Introduction

Pronunciation was once regarded as a "peripheral" and "unteachable" component in ESL/EFL teaching and learning, especially in the trending of "meaning-focused", "communicative" context, which is an area of SLA research that lacks the marriage between theory and practice (Derwing,

2008). However, when communicating with native speakers, learners who fail to discern phonemes of meaningful significance may result in misinterpretation of the message from the interlocutor and thus end up with increased learner anxiety or lack of confidence. Therefore, L1 influences were often identified as sources of these cases. As effective aids to predict L2 sound acquisition issues under L1 influence, speech learning models such as Best (1995) and Flege (2003) respectively used the perceived L1-L2 distance as an indicator of learner performance.

In the Perceptual Assimilation Model (PAM), Best (1995) and Best & Tyler (2007) denote that three native to non-native assimilation types (TC, CG and SC) are discernible when a learner mentally compare two related incoming L2 categories with our native categorie(s) in storage. The L2 sounds may be assimilated to two different L1 sounds, which is the Two Category (TC) type, or to a single L1 category, the Single Category type (SC), or alternatively to a single native category with one being a better candidate than the other, the Category Goodness type (CG). PAM's postulations also include predictions of levels of learners' difficulties in comprehending L2 sounds. The easiest is the TC, then CG, and the hardest one being SC. On the other hand, SLM posits that speakers' L1 and L2 sound systems interact and exist in a common phonological space (Flege, 2003). 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. Too close similarity actually blocks the formation of new L2 categories (Flege, 2003, Flege et al., 2003)

The Mandarin Chinese language has a rich inventory of fricative and affricate consonants ("z/ts/" "c/ts^h/" "s/s/" "j/tɕ/", "q/tɕ^h", "x/ɣ/", "zh/tʂ/", "ch/tʂ^h", "sh/ʃ/", "r/z/") that is more densely categorized than that of English (/f, v, s, z, ʃ, ʒ, ʧ, ʤ/). Specifically, all English fricatives are laminal sounds, i.e., articulation with the tongue tip pointing downwards, but Chinese has a distinction of laminal and apical (upward pointing of tongue tip) sounds.

Wang and Chen (2019) did an experimental study on non-native speakers' perception of Chinese L2 sounds by elementary and intermediate learners, in which they addressed and affirmed the robustness of PAM. The study has explored what English substitutes (among /s, ʃ, ʧ, ʤ, ʒ, t, z, r/) English learners of Chinese resort to when they have to choose one to label Chinese fricatives and affricates audios being played. Their findings suggest that non-native speakers' identification of sounds like "q/tɕ^h/" and "c/ts^h/" are often two- or three-folds, suggesting complicated assimilation patterns. However, their study did not further inquire into whether phonetic contexts play a role in L2 perception.

As pointed out in Lan (2013), certain vowel context can trigger co-articulation in L1 and L2 production, and such gestural economy may also in turn covertly result in varied L2 perception accuracies. Therefore, intending to add original contribution to current literature, this study is especially interested in the impact of vowel context may have on consonant perception.

On a slightly different note, the study is original in two more aspects. In the previous literature, the assimilation types of English fricatives and affricates were experimented mostly on naive listeners. Few studies up to date has given a detailed account of the actual assimilation types as well as learning outcomes of advanced Chinese learners of English who have already received considerably abundant numbers of input. Moreover, previous literature has reported that English learners have problems in distinguishing these phonetic affricates /tr, dr, ts/ from other real phonemic affricates /tʃ, dʒ/ (Lee, 2003; Cruttenden, 2014). In this study, such phonetic affricates (/tr, dr/) were included to find if such confusion exist for Chinese learners.

All being considered, this study aims to tackle the following three research questions based on the

previous findings. The study is believed to be significant in finding the following:

- Q1. How do Mandarin Chinese advanced learners of English categorize English fricatives and affricates (including phonetic affricates) based on Mandarin Chinese candidates?
- Q2. Do vowel contexts of stimuli interact with the consonant perception?
- Q3. Do findings align well with the PAM predictions of category goodness? If not, how can it be explained?

2 Methods

2.1 Participants

Participants are 67 students in 2 parallel English classes the author was teaching in the fall semester of 2019/20. They were all majoring in biomedical sciences, attending a top-ranking provincial university in Guangdong Province, China. Their first language vary in Mandarin (56 cases) and Cantonese (11 cases), but they all use Mandarin daily in their boarding school-life. The average period of formal EFL instruction was around 10-12 years when the experiment took place, and none of the participants and their parents had lived abroad. Their English scores on the National College Entrance Examination ranged between 114 and 138 (out of 150), with an average of 120. Considering their test results and duration of formal English learning, the participants could be collectively described as upper intermediate to advanced learners of English. Prior to the experiment, they have all gone through a 30-minute English IPA training so that they can identify the proper IPAs for the English affricates.

2.2 Stimuli

The stimuli material of the identification task are English fricatives and affricates produced by a professionally-trained, 30 years old, male native American English phonetician, as AmE is the most received variety of pronunciation in that specific area in China, and most participants followed AmE as the language model in their pre-tertiary English-education materials. The target stimuli used in the experiment are monosyllabic English words in CVC structure with /s, ʃ, ʒ, ʧ, ʤ, tr, dr/ as initials except for /ʒ/. Since its lack of onset instances, two

CV₃VC structured words were used instead. Each target phoneme contains variations of three stimuli words, respectively in three extreme vowel contexts /i, æ(a), u/. When a frequent word does not exist under a certain vowel condition, we have increased the number in other vowel contexts to make up the sum of 3. Apart from target stimuli, control words also accompany the target words. We have used plosives irrelevant to the current inquiry (/t, d/) as control fillers, each with three instances, too. The complete list of stimuli can be found in the following table:

| Con-sonant | /i/ | /æ/ | /u/ | 2 nd syllable |
|------------|---------------|-------|-------------|--------------------------|
| s | seed | sack | sued | |
| ʃ | sheep | shack | | |
| tʃ | cheap | chat | chewed | |
| dʒ | jeep | Jack | Jude | |
| tr | | track | truth troop | |
| dr | dream | drag | drew | |
| ʒ | genre Jean | | | visual |

Table 1: Stimuli words grouped by consonants and vowels.

2.3 Procedure

Recording of perception experimental material was prepared 1 month prior to the study. During the recording phase, the AmE speaker was asked to read aloud the stimuli in front of a MD recorder in a sound booth. The recording sampling rate was set at 44100 Hz in mono channel. The recorded stimuli were put in a carrier sentence "Now I say _____", which was also produced by the AmE speaker. The recordings were edited and composed on a PC computer using Praat (Boersma and Weenink, 2020). The target words were separated from the sentences using waveform editing, normalized for peak volume, and saved as wave form for presentations. The stimuli were arranged in random order. The students were given an ISI of 8 seconds after the stimuli to identify the onset of the given syllable of that English word by selecting one of Chinese pinyin "z/ts/" "c/ts^h/" "s/s/" "j/te/", "q/te^h", "x/ε/", "zh/tʂ/", "ch/tʂ^h", "sh/ʂ/", "y/j/", "r/z/".

In the identification experiment, the students are required to listen to the audio presentations designed as above, and press on a button on their cellphone representing those pinyin choices

through an online instant-respond survey system (www.wjx.cn). All participants were then asked to rate the goodness of the English sounds with regard to Mandarin. The goodness was represented in a Likert scale of 0 (very poor) - 5 (exactly the same) on the same system.

One methodological specificity on the presentation of stimuli in tasks have taken controversy: shall we use IPA or Romanization in the presentation of token choices? 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. Especially, Mainland Chinese students often learn *pinyin*, a Latinate Romanization for the language, before the Chinese writing system, hence opening to the possibility that the coincidences of orthography of Chinese and English may play a part in their confusion of L2 sounds in actual acquisition. Therefore, we have utilized pinyin in all L1 Mandarin identification force-choice tasks.

3 Results

3.1 General findings

Overall, the Mandarin speakers had shown organic and yet varied results of cross-linguistic identification. Figure 1 below lays out a straightforward quantitative pattern of participants' choice of sounds, showing a general tendency of identifications patterns over sibilants, affricates and /ʒ/.

The fricative sounds /s, ʃ, ʒ/ behaved differently in the learner's perceptual categorization. /ʃ/ only map on /f/ with a good 4.3 rating out of 5 and 75% categorization. /s/ on the other hand can be mapped both on "s" and "x", with "s" being a better candidate. The situation of /ʒ/ is more complicated with three divided identifications, with the best candidate /r/ only taking up 34.2% of all identifications with the rating fitness at 3.5. The rest two candidates were "y", with 14.7% identification and "zh" at 11.8%, both accompanying low fitness rates of less than 3. More cross-affricate confusion were exposed in the identification of affricates /dʒ, dr, tʃ, tr/. Both /dʒ/ and /dr/ can be mapped on to j and zh with varying percentages from 22.9 to 55.4, with moderate levels of fitness ranging from 3.3 to 4.3. What worth noticing is that the assimilation of /dʒ/ favors "j" and that of /dr/ favors "zh". As for /tʃ/ and /tr/,

both are mostly mapped onto “ch” at over 70%, and the fitness ratings were at a high level of 4.4 and 4.1. /tʃ/ can be identified with an alternative mapping candidate “q”, but with much less percentage, only 25%, but participants who have chosen it showed a high fitness rate at 4.2.

Finally, participants’ perception of control sounds has witnessed over 96% correct categorizations of “t” and “d”. The English plosives has voice onset times different from those of Mandarin but they have been categorized as TC assimilation. Therefore, Mandarin listeners of English naturally map aspirated affricates onto voiceless ones; and unaspirated onto voiced ones.

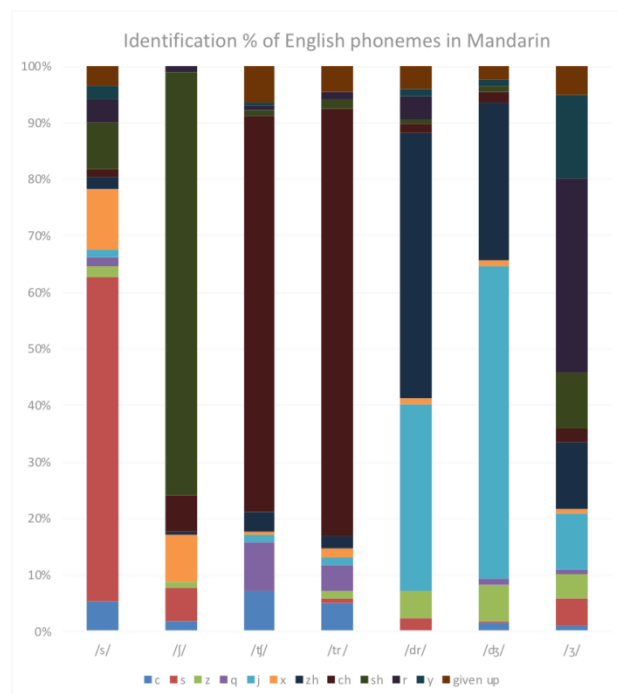


Figure 1: The % identifications of English consonants as each Chinese consonant, stacked to 100%.

3.2 Variation across vowel contexts

Surprisingly, distinctive perceptual assimilation patterns were seen across vowels within each consonant condition. The following Figure 3 demonstrates the frequency of choice of Mandarin candidates in different consonant (pairs) under vowel contexts of /i, a, u/. The order of the presentation is /s, ʃ/, /tʃ, tr/, /dʒ, dr/ and /z/ from the upper to the lower panels. In the /s, ʃ/ pair, although the consonants favored “s” and “sh” respectively in general (both at more than 80%),

the choice of other stand-alone candidates across vowels showed variation. As for /s/, 13 out of 134 outlier cases, a significant higher rate of choosing “sh” in the /u/ context than 1-4 cases in other two contexts can be identified, whereas for /ʃ/, only consonants in words with vowel /i/ are prone to be categorized as “x”, at 14 out of 170. A reverse pattern can be identified.

The vowel-context variation for /tʃ/ showed that the sounds with /i/ behaves slightly different from those with the other vowels: /æ(a)/ and /u/ both had fewer than 3 cases categorized as “q” but /i/ had a considerable 13 cases as “q”. As for “tr” however, almost all sounds were mapped onto “ch” regardless of the vowel context, all with exceptions of as few as 3 instances of categorizations as “q”.

The /dʒ, dr/ pair showed a distinct vowel context variation that worth noticing. In terms of general tendency, participants associate /dʒ/ with “j” and /dr/ with “zh”. Though both sounds can be categorized as “zh” and “j”, /dʒ/ in the /u/ context showed a reversed pattern against the other two contexts, favoring “zh” at 35 cases of “zh” to 23 cases of “j”, whereas /dr/ in the /i/ context behaved opposite from /tr/, favoring “j” at 28 cases of “zh” to 33 cases of “j”.

The behavior of the consonant /z/ is the most complicated across vowels within these four panels of graphs. Participants’ categorization of /z/ as “r” exists in all three conditions, and topped in /i/ at 49 cases, but significantly fewer in /u/ conditions at only 19 cases. The “zh” choice exist only in /i/ and /æ(a)/ contexts, but not /u/. The third candidate “y” is the favored choice of participants only in the /u/ context, which was labeled 26 instances among the total 45 cases in this vowel context, well over half; but not popular at all in the other two contexts, constituting a distinct pattern. There is a clear reverse pattern between /tr/ and /y/ in the /i/ and /u/ vowel contexts.

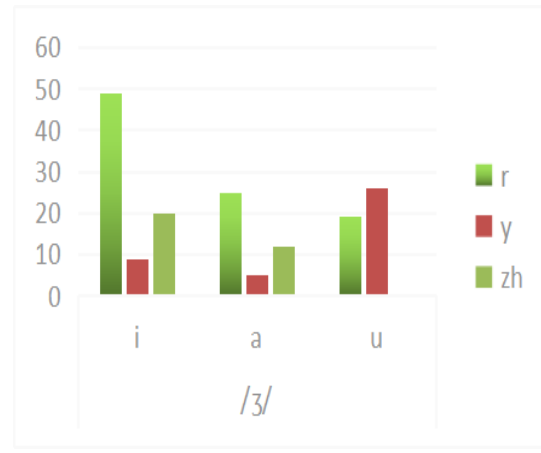
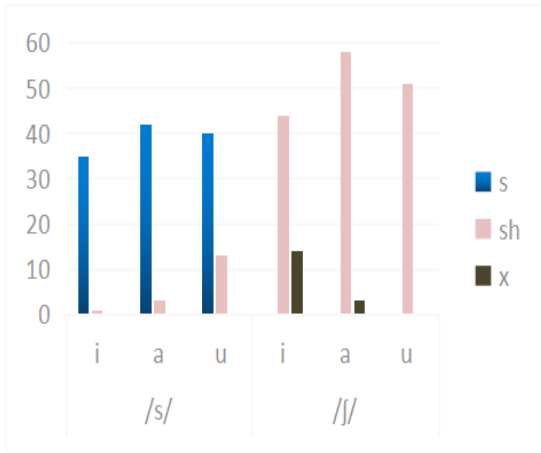
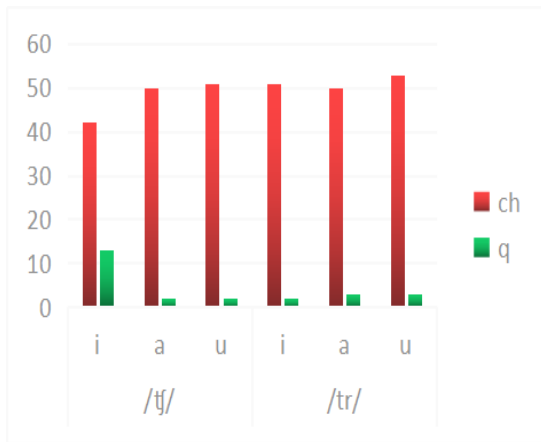


Figure 2: The % correct Mandarin categorizations of English consonants (From top to bottom: /s, ʃ, /ʒ tr/, /ʒ, dr/ and /ʒ/) clustered by vowels.

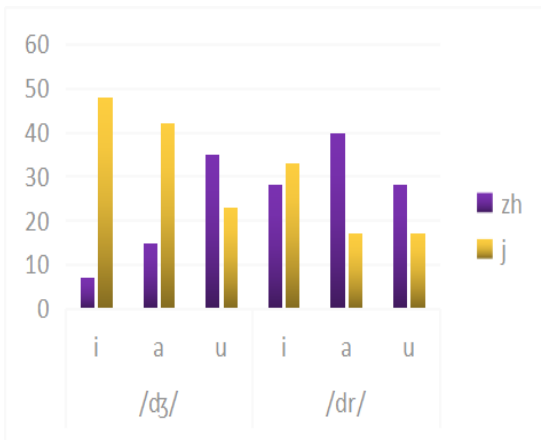


4 Discussions

As predicted, Mandarin candidate consonants were mapped onto English sounds within a range of fitness ratings from 1.0 (poorest) to 6.4 (best) by native Mandarin listeners. Some assimilation patterns elicited from the results do conform to predictions by PAM. For example, English sibilants /s, ʃ/ showed largely similar assimilation tendencies as reported in the previous literature.

However, the variation within categories may not be predicted by PAM. For example, /tʃ/ and /tr/, which supposedly constituted a classic SC (single category) assimilation pattern, showed CG-like patterns. For example, PAM could explain that some participants preferred the “ch” in Chinese as a candidate for /tr/. Since the /r/ sound in English does not exist in Chinese, so this “cluster” is seldom considered as separated sounds by Chinese listeners. So it is reasonable to predict that they would regard “ch” as the equivalent to /tr/. However, some other listeners preferred the candidate “q” but only in limited vowel contexts of /i/ disregarding its similarity to “ch”. This is probably due to the fact that “q” and /tr/ share similar places of articulation in the vowel context of /i/, where the palatal “q” and post-alveolar /tr/ approaches physiologically. This leads to the main argument of this research that vowel context may greatly shift the L1-L2 perceptual space, which questions the idea that such a space may remain intact during a specific learning stage (Flege, 2003).

The variation of vowel context shown in many consonant perceptions could be traced of



physiological roots. For example, the /s/ and /ʃ/ context showing /i/ and /u/ as behaving differently from the other two contexts is clearly affected from the gestural proximity of articulatory gestures. The /s/+/u/ combination will result in palatalization because the upward movement of the tongue dorsum by /u/ can pull backwards the articulation of fricative /s/, creating post-alveolar frication (cf. Lan, 2013). Similarly, /ʃ/+/i/ in the front vowel context will add the degree of frontness of the tongue tip to the post-alveolar sound, pulling the tongue body further front, palatalizing the sound to make it similar to the Chinese “x” which only exists before front vowels /i, y/ phonologically.

Vowel influence can also be seen in /ʃ/, ɟʒ, dr/. The /ʃ/ sound being categorized as “q” and /ɟʒ/ as “j” in the /ʃ/+/i/ or /ɟʒ/+/i/ combination has a phonological constraint of only allowing front vowels /i, y/ after “j” and “q”, which is in line with gestural economy: the palatalization may come from the co-articulation due to increased degree of frontness of the tongue tip (see Browman and Goldstein, 1992; Gick et. al., 2006 for similar examples of /r/ in different vowel contexts). Such a tendency can also be seen grammaticalized in many loanwords such as Jeep (ji pu), Cheetah (qi ta) and so on. As for the /u/ context, /ɟʒ/+/u/ is often regarded as a retroflex affricate in Chinese, which was also noticed as an unexpected finding in Wang and Chen (2019). What they did not specify is that these two sounds share the common place of articulation, and only differs in apicality. Therefore, we can see that the difference between apical and laminal sounds was considered not sensitive by Mandarin speakers of English, who are more sensitive to places of articulation.

The English voiced fricative /ʒ/ is another sound that has caused mass confusion in categorization. It is generally categorized as “r”, or the semi-vowel “y”(j). The distribution of the assimilation shows that “r” is the preferred variant especially in the context of /i/ and “y” mainly preferred in the context of /u/. The reason behind it may lie in that “r” as a fricative sound /ʒ/ shares similar place of articulation (post-alveolar) and only differs from /ʒ/ in terms of apicality. Phonologically, “r” can only exist in the context of a front vowel /i /, which is an allophonic variation and graphic equivalent of /i/ in Mandarin. However, as for /u/, the reason of learners favoring “y” is unclear.

Although /dr/ and /ɟʒ/ are both similar to the “zh” sound in Mandarin, they are not assimilated into a single category SC because an additional CG candidate, “j” also exists in the Mandarin phonemic inventory, with “zh” as the more preferred target (70%) for “dr”. Although Wang and Chen (2019) found something similar, our study showed, interestingly, that /ɟʒ/ has the more similar representation with “zh” when considered phonetically: “zh” is more phonetically similar to /ɟʒ/, which is proven in the very stable categorization of /ʃ/ as “sh”, but not the other candidate “x”, which only differs in voicing (see Table 2). Therefore, we could conclude that the above findings cannot be solely attributed to phonetic similarity claimed by PAM. A possible explanation to this anomaly is the orthographic coincidence of Chinese “j” and English /ɟʒ/ in Romanization. Such orthographic influence will be studied and argued for in another study.

| Mandarin | phonetic difference | % of identification | predicted match |
|----------|-------------------------|---------------------|-------------------|
| /ʃ/ | | | |
| x | place * | 9.5% | poor |
| sh | manner (apicality) | 75% | very good |
| /ɟʒ/ | | | |
| zh | manner (apicality) | 22.9% | poor |
| j | both manner and place * | 55.4% | good to very good |

Table 2: Assimilation mismatches of English /ʃ/ and /ɟʒ/ in terms of phonetic-similarity-predicted match and actual identification proportion.

Finally, the findings reflects on Derwing and Munro (2004) and Major (2014)’s concern that in actual L2 perception, learners may utilize other phonetic details such as duration, or phonetic context, which may win them high confidence in discerning those sounds. But, some of them may not be correct and helpful to learning, which can be seen in various modes of assimilation. As for the SC-type pairs in the current experiment, i.e., English /ʃ/ and /tr/, learners show very high identification as Chinese “ch” (70.1% and 75% respectively), with high goodness ratings. For TC pairs, For example, the /ʃ/ sound has experienced a relatively low accuracy compared to the apparent categorization could be predicted in the L1 Chinese candidate “sh”. However, both

identifications may lead to low accuracy in actual perception, though not attested in the current inquiry. That being said, a good category for L1-L2 mapping may be still perceptually confusing. At this point, we could affirm that the perception of L2 sounds is more than a low-level acoustic process that is merely generated by automatic distance-comparison, but must include higher-level cognitive processes that alter with par linguistic and discourse contexts.

5 Conclusion

The study presents a novel design in an attempt to understand the vowel influence on L2 consonant perception. Both theoretical and pedagogical implications can be drawn from the findings. Much of the assimilation patterns surfaced as predicted by those from TC, CG and SC types in the PAM model. However, on the other hand, the perceived phonetic distance between L1 and L2 is not the only factor in play. We have witness that the assimilation patterns were under the influence from physiological attributes such as gestural economy levels in various vowel contexts.

Theoretically, we could see from the results that vowel contexts clearly interact with L2 phonology. In other words, the universal human speech apparatus enables the phonetic sounds to express in phonology without boundaries, which has affected and complicated the “should-be” holistic interlanguage in phonological terms. The L2 evidence may be seen as a side proof of the view of gestural nature of phonological perception (Liberman and Mattingly, 1985; Browman and Goldstein, 1992).

Pedagogically, the study invites instructors and learners to admire and tackle the complexity of speech learning and the L1 influence, especially in families of sounds like fricatives and affricates of L1 Chinese and L2 English where multiple entangling mappings can be found. Future studies should include perceptual accuracy tests and production tests to further solidify the implications of current findings, especially whether the actual learner accuracy performance was as predicted by theoretic models so that more theoretical and suggestions can be made. The non-linguistic effect, for example orthographic influences, should also

be carefully explored so as to discover more myths in the complex system of L2 consonant perception.

Acknowledgements

This study is financially supported by the Shenzhen University Young Scholar Start-up Fund (No. 2019082).

References

- A. Cruttenden. 2014. *Gimson's Pronunciation of English*. 8th edition, Routledge Tylor & Francis, England.
- A. M. Liberman and I. G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 2 (1): 1–36.
- B. Gick, F. Campbell, S. Oh, & L. Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, 34, 49-72.
- C. P. Browman and L. Goldstein. 1992. Articulatory phonology: an overview. *Phonetica*, 49, 155-180.
- C. T. Best. 1995. A direct-realist view of cross-language speech perception,” *Speech perception and linguistic experience: Issues in cross-language research*, 171-204.
- C. T. Best and M. Tyler. 2007. Nonnative and second-language speech perception: commonalities and complementarities. In O.S. Bohn & M. Munro (Eds.). *Second-language Speech Learning: the Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James. E. Flege*. Amsterdam: John Benjamins, 13-34.
- J. E. Flege. 2003. Assessing constraints on second-language segmental production and perception,” *Phonetics and phonology in language comprehension and production: Differences and similarities*, 319–355.
- J. E. Flege, C., Schirru., and I. R. A. MacKay. 2003. Interaction between the native and second language phonetic subsystems, *Speech Communication*, 40, 467-491.
- P. Boersma and D. Weenink. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.09, retrieved 26 January 2020 from <http://www.praat.org/>
- R. C. Major. 2014. *Foreign Accent—the ontogeny and phylogeny of second language phonology*. New York: Routledge.

- S. Lee. 2003. A comparison of cluster realizations in first and second language. *The Journal of Studies in Language*, 19(2): 341-357.
- T. Derwing and M. J. Munro. 2005. Second Language Accent and Pronunciation Teaching: A Research - Based Approach. *TESOL Quarterly* 39 (3): 379-397.
- T. Derwing. 2008. Curriculum issues in teaching pronunciation to second language learners. In J. Hansen Edwards & M. Zampini (Eds.). *Phonology and Second Language Acquisition*. Amsterdam, Netherlands: John Benjamins, 347-369.
- X. Wang and J. Chen. 2019. English Speakers' Perception of Mandarin Consonants: The Effect of Phonetic Distances and L2 Experience. *ICPhS 2019 Proceedings*, 1-5.
- Y. Lan. 2013. Towards a Revised Motor Theory of L2 Speech Perception, *Proceedings of PACLIC 27*, 136-142.