INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Perception of English Fricatives and Affricates by Advanced Chinese learners of English

Yizhou Lan

School of Foreign Languages, Shenzhen University, China

yzlan@szu.edu.com

Abstract

70 Mandarin-speaking advanced learners of English (level B2 and above) participated in a perceptual identification experiment eliciting their preferred Mandarin equivalent classifications of English fricatives and affricates (/s, ʃ ʧ, ʤ, tr, dr, ʒ/) along with fitness rates. The degree of mapping between Mandarin and English consonants, ranging from poor to fair, and good, were compared against predictions by the Perceptual Learning 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. The Findings suggest that phonetic similarity do predict most mappings as the learning models postulate, but other factors such as articulatory proximity and orthographic influences should be considered, too.

Index Terms: L2 speech acquisition, fricatives, affricates, consonant cluster, Perceptual Assimilation Model

1. Introduction

Previous endeavors of understanding L2 speech acquisition such as Best [1] and Flege [2] who had respectively put forward the perceptual assimilation model (PAM) and speech learning model (SLM), describe the process of L2 learning over time as a function of perceived L1-L2 phonetic distance. In the PAM model, 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. 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 [2] [3].

Chinese phonology has a rich inventory of fricatives and affricates "z/ts/" "c/tsʰ/" "s/s/" "j/tɕ/", "q/tɕʰ/", "x/ɕ/", "zh/tʂ/", "ch/tʂʰ/", "sh/ʂ/,", "r/ʐ/", which include both voiceless and voiced sounds in alveolar, post-alveolar, retroflex and palatal places of articulation. Such rich density of phonemic categories in the L1-L2 common perceptual space is believed to be beneficial in learning a language with fewer similar categories [4]. However, no recent study 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.

With regard to the methodologies of assessing phonetic distances between L1 and L2 speech sounds under our current inquiry, using acoustic properties of phones across languages may not be sufficient as such measurement may not capture the most crucial phonetic cues of category formation [5]. The commonly used method of phoneme inventory comparisons as the perceptual element in tests, the IPA, is not applicable either as the phonetic symbols do not provide the detailed phonetic properties of sounds for Chinese learners of English. Mainland students, from the beginning of schooling, acquire Chinese through 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.

Wang and Chen [5] reported a series of state-of-the-art results on non-native speakers' perception of Chinese L2 sounds by elementary and intermediate learners, which has inspired the study as a reverse replication. The study has identified the L1 English substitutes English learners of Chinese resort to as they face a force-choice task on naming Chinese fricatives and affricates. The examined L2 English candidates were namely /s, ʃ, ʧ, ʤ, ʒ, t, z, r/. Their findings suggest that non-native speakers' identification of sounds like "q/tɕʰ/" and "c/tsʰ/" are often two- or three-folds, suggesting complicated assimilation patterns. However, they have not identified sounds that can be phonemically cluster but phonetically similar to the sounding of some Chinese candidates, such as "ch" and "q".

Several studies [6] [7] [8] have categorized some English consonant clusters (/tr, dr, ts/) as phonetic affricates in the L2 speech and ESL context. According to a pilot survey of Chinese students' self-reported pronunciation problems, they may confuse these phonetic affricates from other real phonemic affricates /ʧ, ʤ/ in terms of both perception and articulation, thus complicating the issue of English affricate/fricative perception. In this study, such phonetic substitutes were included to find if such confusion exist for Chinese learners.

The study is thus interested in finding the mapping patterns of L1 Mandarin sounds on L2 English ones, and the linguistic and non-linguistic motivation behind them, especially when using the pinyin system as the percept for assimilation, with the phonetic affricates added to the test material.

http://dx.doi.org/10.21437/Interspeech.2020-1120

2. Method

2.1. Participants

70 students in 2 parallel English classes participated in the two experiments with a 2-week interval in between. Participants' gender ratio of male to female students is 5:3. They are all Mandarin-speaking college students with 11 cases also speaking Cantonese. Their English scores on the National College Entrance Examination ranged between 108 and 140 out of 150 with an average of 120, which 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 and Procedure

Stimuli materials for the experiments are English fricatives and affricates produced by a trained native English speaker and phonetician, in standard American English. The stimuli are mainly CVC monosyllabic words with selected fricative, affricate and phonetic affricate initials /s, ʃ ɹ, ʒ, ʧ, ʤ, tr, dr/. As for /ʒ/. On top of that, two CVʒVC structured words were also included in the stimuli. Each target phoneme was introduced by three stimuli words, respectively in three extreme vowel contexts /i/, /æ/(/a/) and /u/. We have added irrelevant fillers with /t, d/ as initials for controlling the variants. The complete list of stimuli can be found in the following table:

Table 1: *The stimuli chart columned by consonant.*

	/i/	/æ/	/u/	2nd syllable	fillers
s	seed	sack	sued		
ʃ	sheep	shack			
ʧ	cheap	chat	chewed		
ʤ	jeep	Jack	Jude		
tr		track	truth troop		
dr	dream	drag	drew		
ʒ	jean	genre		visual	
t					tea tag tooth
d					deep deck

The native English speaker was asked to read aloud the stimuli in a carrier sentence "Now I say ______" in front of a MD recorder in a sound booth. The recordings were then edited on a PC computer: the target words were separated from the sentences using waveform editing through Praat [9]. The target words were normalized for peak volume and duration, and put in a uniform version of the carrier sentence recording, and finally saved as .wav for presentations.

In the first experiment, the students were required to listen to the recorded tasks in random order. Participants were given an ISI of 8 seconds after the stimuli to identify the onset of the given syllable of that word from 11 Chinese pinyin choices "z/ts/" "c/ts^h/" "s/s/" "j/tɕ/", "q/$tɕ^h$/", "x/ɕ/", "zh/tʂ/", "ch/$tʂ^h$/", "sh/ʂ/,","r/ʐ/", and "y/j/". An additional option of "I give up" was added. Along with the 12 choices of identification, all participants were asked to rate the goodness of the chosen Chinese candidates with regard to the English stimuli. The goodness was represented in a Likert scale of 0(very poor)-5 (same). The total number of tokens in experiment 1 is 70 for each participant and the participant responses were collected from an online real-time survey system (www.wjx.cn) during a meeting of the class.

The second experiment, a perceptual accuracy test, took place in the same venue during a class 2 weeks after the first experiment. Participants were given two sub-sections of tasks. In the first half, stimuli in carrier sentences, exactly the same as in the first experiment, were given to students through a speaker. Participants were required to choose from the online survey system the correct IPA from the recordings after an ISI of 5 seconds for each token. The second half of Experiment 2 was a force-choice same/different discrimination task. Target pairs in carrier sentences containing two sounds, either differing in the onset fricative/affricate or being completely the same, were presented to the participants with an ISI of 5 seconds. The participants need to choose "same" or "different" on the online survey system indicating either the two stimuli are same or different during that period. The total number of tokens for experiment 2 is 70 for each participant.

3. Results

3.1. Experiment 1

The mean percentages of identifications of the 7 English fricatives and affricates as Mandarin sounds are presented in the following Table 1, along with the rating scores and match predictions. As for the percentages of identifications, the identification rates of a given sound are presented in average across all participants. The matching types are also provided based on the criteria of proportionate ratings par identification (= percentage × goodness ratings). The "Good", "Fair", and "Poor" matches were labeled with standards of "≧3; 1.5-3; <1.5" respectively.

Table 2: *Learner identification percentage and fitness ratings for English phonemes.*

English phonemes	Mandarin Identification (pinyin)	% identified	rating	match
ʧ	ch	70.1	4.4	Good
	q	25	4.2	Fair
tr	ch	75.7	4.1	Good
dr	j	32.8	3.3	Fair
	zh	47.1	4.0	Good
ʒ	r	34.2	3.5	Fair
	y	14.7	2.8	Poor
	zh	11.8	2.6	Poor
s	s	57.4	4.3	Fair
	x	10.8	4.1	Poor
ʃ	sh	75	4.3	Good
ʤ	j	55.4	4.3	Fair
	zh	22.9	3.5	Poor

As shown in the table, participants produced complicated mapping patterns of cross-linguistic identification. Though clear preferences for the assimilation candidate of each stimulus can be seen over others, they were not all simple one-on-one mapping categories. Five of the seven English consonants had the modal classifications (ones that take the largest percentages) over 50%, some as high as 75%, but others were as low as 20%. For fricatives, /s, ʃ, ʒ/ behaved differently as /ʃ/ only map on "sh" 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 at 57.4% in terms of categorization. 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 for /ʒ/ 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 /ʤ, dr, ʧ, tr/. Both /ʤ/ 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 percentage for /ʤ/ favors "j" and that for /dr/ favors "zh". As for /ʧ/ and /tr/, both are mostly mapped onto ch at 70.1 and 75.7 respectively, and the fitness ratings were at a high level of 4.4 and 4.1. /ʧ/ has an alternative mapping candidate "q", but with much less popularity of only 25% together with a surprisingly high fitness rate at 4.2. Figure 1 below shows the identified assimilation patterns with the dashed boxes indicating English L1 consonants and solid boxes as Mandarin L2.

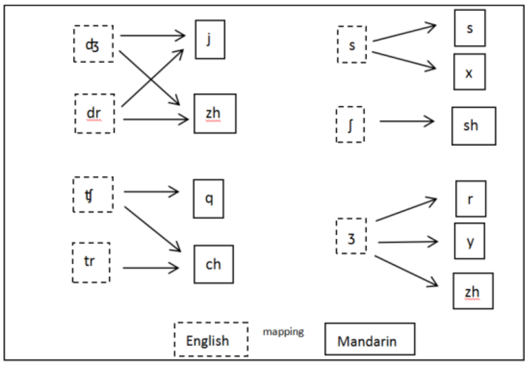

Figure 1: *Assimilation patterns of English fricatives and affricates onto Chinese, identified by learners.*

3.2. Experiment 2

In the second experiment participants have performed in the ABX discrimination test with varied accuracy rates across target consonant pairs. The percentages for correct responses are 55.71% and 71.90 for /ʧ/ and /tr/, 79.52% and 68.1% for /s/ and /ʃ/, 63.33% for /dr/ and 67.14% for /ʤ/, and 75.95% for /ʒ/, showing a considerably wide range of accuracy.

The same/different discrimination tasks showed similar results of the above. The overall accuracy of the discrimination tasks is 78.8%. However, considering the nature of the task, we have divided the data of the task into two folds: items with same tokens, where participants listen to and may get true positive or false negative results; and items with different tokens, where participants may produce false positive or true negative results. As for the items with same tokens, the overall true positive rate is high as 89.9%, with the accuracy rates of tokens concerning /tr/ &/ʧ/ at 88.8%; those of /dr/ & /ʤ/ at 90.9%, and /ʒ/, /r/, and /ʤ/ at 88.4%. The overall true negative rate for different tokens, however, was as low as 66.5%. The accuracy rates of discriminating different /tr/-/ʧ/ pairs was 50.0% at chance level, /dr/- /ʤ/ pair at 75.4 and /ʒ/-/r/-/ʤ/ set at 68.1%. The following table lays out the above accuracy rates aligned with PAM predictions.

Table 3: *Accuracy rates and PAM predictions by task question pairs in the discrimination task.*

	Task pairs	PAM type	Accuracy rate	D/S rate	Overall rate
DIFF tokens	/tr/-/ʧ/	SC	50.0%	66.5%	78.8%
	/dr/-/ʤ/	SC	75.4%		
	/ʒ/-/r/	CG	68.1%		
SAME tokens	/tr/-/ʧ/	SC	88.4%	89.9%	
	/dr/-/ʤ/	SC	88.8%		
	/ʒ/-/r/	CG	90.9%		

4. General Discussions

As shown in Figure 1, L1-L2 mappings derived from Experiment 1 were complicated, with each sound having 1-3 candidates, whose fitness ratings range from 1.0 (poorest) to 6.4 (best) by native Mandarin listeners. Many assimilation patterns do conform to theoretical predictions by learning models. As predicted, TC categories were perceived with congruent model classifications. The control sounds has witnessed over 96% categorization to "t"and "d". The sound has voice onset time of Mandarin and English are different but they are categorized as TC assimilation. Therefore, Mandarin perceivers of English naturally map aspirated affricates onto voiceless ones; and unaspirated onto voiced ones.

While most mappings can be predicted by L1-L2 phonetic distance as postulated by PAM, as in the previous reverse study [5], the identification of some targets did not behave as predicted. For example, according to PAM, /s/ should be assimilated to "s" since the acoustic and perceptual details of English /s/ and Mandarin "s" are almost identical, disregarding phonological specifics. However, the sound /s/ has only 57.4% categorizations of "s", with 10.8% exception of "x". This is different from the PAM prediction that "s" being a good candidate of /s/. Inquiring L1 Mandarin phonology, the emergence of "x" may be attributed to the vowel context of /i/ that triggers palatalization of the sound /s/, resulting in the resemblance to "x", a /i/-specific allophonic variation of "s" in Mandarin. Likewise, participants preferred "q" to be assimilated to /ʧ/ in limited vowel contexts of /i/, again probably because /ʧ/ and "q" share the palatal or post-alveolar position. Therefore, vowel context is another factor that shapes learner perception.

Secondly, an intriguing finding is that consonant clusters can be an equally competent candidates in L2 perception of single consonants. For example, as PAM predicts, /ʧ/ can be a good English equivalent for "ch", but "tr" being a consonant cluster which generates syllabification different from /ʧ/, should not be a good candidate for "ch". Therefore, /tr/&/ʧ/ - "ch" should be a CG type of categorization favoring /ʧ/, with fair to good accuracy [1]. However, results in experiment presented that /ʧ/ is 72% and /tr/ is 80% categorized as "ch" with /tr/ slightly higher, which has constituted a SC assimilation type, resulting a 50% low accuracy in the same/difference task. Since the /r/ sound in English does not exist in Chinese, and the consonant cluster is not a legal syllabic configuration in the Chinese syllabic structure (C)V(N), the cluster "tr" will be considered a single affricate perceptual candidate, affirming Cruttenden's classification [8] that /tr/ is a single sound from the advanced learner's point of view.

A closer look at the /tr/&/ʧ/ -"ch" mappings showed more insights. These two contrasts yielded high rates of modal classification (70% & 71% for /tr/ and /ʧ/ respectively), both fitting well with "ch", but incompatible with low accuracy rates in the same/difference task accuracy results, favoring /tr/ over /ʧ/. This implies that in actual L2 perception, learners may utilize other phonetic details such as duration, or phonetic context, that may help them discern the sounds, though some of them may not be helpful.

The voiced affricates showed as much complication. /dr/ has two possible assimilation directions: "zh" and "j", with "zh" as the preferred target (70%), whereas /ʤ/ is also assimilated to these two same Chinese sounds, but with "j" as the preferred target. This conforms to PAM predictions that these two sounds should be CG types (Fair matches) of assimilation. However, there is a mismatch in the phonetic similarity and perceptual similarity in these two preferred choices. For example, "zh" instead of "j" is phonetically more similar to /ʤ/. A possible explanation to this anomaly is the orthographic (mis)representation of the spelling of "j" in English has led to the pronunciation of /ʤ/ when surfaced as j similar to the pinyin grapheme "j", as the loanwords in Chinese originated from "Jack, Jay, Jeep" were transliterated in pinyin as "jieke, jie, jipu". Such a grapheme-phoneme connection was then established and entered the participants' perceptual realm.

Orthographic influence from Romanization coincidences also explains the categorization of /ʒ/ as "zh". Both /ʒ/ and /ʤ/ could be mapped onto the Chinese "zh", suggesting that Chinese students are affected by orthographic confusions of "g" that sounds /ʒ/ as in "genre" with its counterpart /ʤ/ of relatively higher functional load. Learners have shown difficulties in identifying it from the /ʤ/ as in "germ, general, George, judge, etc". Similar to the pinyin confusion, the English irregular spellings has given L2 learners another challenge in speech perception. These finding poses questions to PAM suggesting that the perception of L2 sounds is more than an low-level acoustic process that is merely generated by automatic distance-comparison, but a higher-level cognitive one that is warped by graphic, and probably other multi-modal contexts [3] [10].

In Experiment 2, the same/different discrimination results have borne out accuracy rates aligning well with PAM predictions, showing that SC patterns are significantly smaller than CG patterns. For example, SC patterns such as /tr/-/ʧ/ and /ʤ/-/dr/ are perceived with low levels of accuracy near chance level. But the big gap between false positive errors and true negative errors shows that participants are good at identifying similarities but encountering trouble in discerning differences, suggesting that learners would neglect certain important phonetic cues due to inaccurate perception of phonetic distance. It may suggest that when the students are performing complex tasks that involve comparison of multiple sounds, such as ABX identification, the accuracy may be lowered due to heavy cognitive load within a relatively short ISI [3].

5. Conclusion

The study possesses a number of limitations regarding experiment design. Inherently it is difficult to strike a balance between the force-choice behavioral experiments and the free expression of their reflections on pronunciation difficulties. During the experiments, the participants have performed extraordinarily in completing the very tedious and almost confusing experiments with their good intentions and resilience. The accuracy differences across tasks need further investigation on the degree to which cognitive load these tasks poses on experiment subjects.

In conclusion, the present study gives both theoretical implications on the predictability of phonetic distance on L2 learning outcomes. In both experiments, especially experiment 2, the identification and discrimination tasks produced similar learning outcomes by L2 Chinese learners as predicted by those from TC, CG and SC types of English-Chinese consonant mappings as postulated by the PAM model. Results have clearly shown that even advanced learners, after years of L2 input and communication, are no exceptions of L1-induced learning difficulties rooted in their linguistic landscape and the perceived distance of L1-L2.

However, on the other hand, the present study showed that the perceived phonetic distance between L1 and L2 is not the only factor in play, and learning outcomes are under the probable influence from L2-L1 orthographic coincidences may give rise to unexpected patterns of assimilation. Future studies should include production tests for the comparison against identification and discrimination results so that more production-related theoretical models such as SLM can be further attested.

6. Acknowledgements

This paper is supported by Research Foundation of Education Bureau of Guangdong Province, China (2018GXJK165), Shenzhen University Young Scholar Start-up Funds (No. 2019082) and the Sapiens Language Research Center, School of Foreign Languages, Shenzhen University.

7. References

[1] C. T. Best, "A direct-realist view of cross-language speech perception," *Speech perception and linguistic experience: Issues in cross-language research*, pp. 171–204, 1995.

[2] J. E. Flege, "Assessing constraints on second-language segmental production and perception," *Phonetics and phonology in language comprehension and production: Differences and similarities*, pp. 319–355. 2003.

[3] W. Strange, "Automatic selective perception (asp) of first and second language speech: A working model," *Journal of Phonetics*, vol. 39, no. 4, pp. 456–466, 2011.

[4] R. C. Major, *Foreign accent: the ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates. 2001.

[5] X. Wang, and J. Chen, "English Speakers' Perception of Mandarin Consonants: The Effect of Phonetic Distances and L2 Experience," *ICPhS 2019 Proceedings*, pp. 1-5, 2019.

[6] S. Lee, "A comparison of cluster realizations in first and second language," *The Journal of Studies in Language*, vol. 19, no. 2, pp. 341–357, 2003.

[7] F. R. Eckman, "The structural conformity hypothesis and the acquisition of consonant clusters in the inter-language of ESL learners," *Studies in Second Language Acquisition*, vol. 13, no. 1, pp. 23–41, 1991.

[8] A. Cruttenden, *Gimson's Pronunciation of English, 8th edition*. Routledge Tylor & Francis, England, 2014.

[9] P. Boersma, and D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 6.1.09, retrieved 26 January 2020 from http://www.praat.org/

[10] D. Larsen-Freeman, "Chaos/Complexity science in second language acquisition", *Applied Linguistics*, vol. 8, 1997.