The Perception of English Vowel Contrasts by Chinese EFL Learners and Native English Speakers

Aihui Zhang[1], Hui Feng[1,2,3], Jianwu Dang[3, 4], Xinyuan Zheng[3] and Zhihao Xu[5]

[1]School of Liberal Arts and Law, Tianjin University, China
[2]Research Center for Linguistic Sciences, Tianjin University, China
[3]Tianjin Key Laboratory of Cognitive Computing and Application, China
[4]School of Information Science, Japan Advanced Institute of Science and Technology, Japan
[5]Nankai High School, Tianjin, China
zhangaihui111@126.com, fenghui@tju.edu.cn, jdang@jaist.ac.jp

ABSTRACT

Previous studies have shown that some English vowel contrasts remain difficult for English learners to perceive. This study investigated three potential factors affecting vowel perception: cross-language vowel systems, spectral distance, and language proficiency, for the discrimination and identification of English /ɪ/-/ɛ/, /ɛ/-/æ/, /ʊ/-/ʌ/, and /ʌ/-/ɒ/ contrasts in four groups of listeners: L1-Chinese with beginner level, low proficiency, and high proficiency in English, and L1-native English speakers. Eight 10-step vowel continua were used in identification and AXB discrimination tasks. Results reveal that the disparity in the correctness in the discrimination task for L1-Chinese is higher than that for L1-native English speakers. No overall positive correlation is found between spectral distance and vowel discrimination, though positive correlation can be found in either front or back vowel contrasts when discussed separately. As language proficiency is enhanced, front vowels are perceived more categorically, while such a tendency is not shown in back vowel contrasts.

Keywords: cross-language vowel systems, spectral distance, language proficiency, vowel perception

1. INTRODUCTION

Previous studies have shown that Chinese EFL(English as a Foreign Language) learners have difficulties distinguishing some vowel contrasts. Ho [9] showed that Chinese EFL learners perceived /æ/ better than /e/ and /ɛ/. Tseng [16] found that mal-perception existed among /e/-/ɛ/-/æ/, with /ɛ/ the hardest and /æ/ the easiest for Taiwan EFL learners to perceive. Lin [11] proved that mainland Chinese EFL learners performed worst in /e/-/ɛ/-/æ/ perception, where /e/ was misperceived as /ɛ /, and /æ / as /e/.

Acoustic cues, such as differences of cross-language vowel systems, spectral distance of L2 vowels, and durational cues have been discussed as factors contributing to the difficulty in L2 vowel perception. Studies of cross-language speech systems show that L2 learners are insensitive to certain phonemes which are realized as different allophones in L1 but are realized as independent phonemes in L2. For example, Chinese EFL learners may mingle vowel contrast /ɛ/-/æ/ as /ɛ/-/æ/ are allophones of /a/ in Chinese vowel system [17]. Escudero, et al found that both L1-Dutch and L1-German listeners weight vowel spectrum heavier than vowel duration, whereas L1-Spanish L2-Dutch listeners favor vowel duration [6]. Mora, et al discovered that native Spanish adults have difficulties perceiving Catalan mid-vowels /e/-/ɛ/ contrasts due to small discrepancy between adjacent vowels [14]. Duration is also found as an important cue for German English learners to distinguish the English vowel contrast /æ/-/ɛ/ [3].

In the case of Chinese EFL learners, whose vowel system differs from that of English in that Chinese has no /i-ɪ/, /u-ʊ/, /ɛ/-/æ/, or /ʌ/-/ɒ/ contrasts [17]), perception experiment of cross-language family languages by Chinese EFL learners may present a different picture from studies which focus on learners from the same language family [eg. 3, 7, 13, 15]. Since vowel duration is not meaning-making in Chinese, spectral distance may play a vital role in the perception of some English vowel contrasts.

Based on previous studies on the factors affecting the perception of English vowel contrasts, this study is designed to test whether and how the two acoustic cues, spectral distance and cross-language vowel systems, affect the perception of English contrasts /ɪ/-/ɛ/, /ɛ/-/æ/, /ʊ/-/ʌ/, and /ʌ/-/ɒ/ by Chinese EFL learners. The role of language proficiency will also be investigated, since a potential relation between vowel perception and language proficiency is revealed in our pilot study.

2. EXPERIMENT

2.1. Subjects

A total of 20 adults were chosen as subjects, who were classified into four groups according to their

English proficiency measured by IELTS score (Table 1). For beginner lever Chinese EFL learners, their IELTS score was not available as they have studied English for no more than 30 hours. The four native English speakers are English teachers in Chinese universities, who are conventionally considered to have the highest proficiency, though they have not taken IELTS tests. Prior to the perception experiment, a short questionnaire of language learning experience was conducted for subsequent screening.

Table 1. Features of Subjects; "C" for Chinese: "E" for English; "—" means "not available".

Group	Features	No.	IELTS score
1	L1-C, L2-E Beginner	7	—
2	L1-C, L2-E Intermediate	4	5.0-6.0
3	L1-C, L2-E Advanced	5	>7.0
4	L1-E, native English speaker	4	—

2.2. Stimuli

"Pure" English vowels with no preceding or following sounds were chosen as stimuli according to Mora [14]. The stimuli used in both identification and discrimination tasks were drawn from 4 synthetic vowel continua: high-mid and mid-low front vowel continua (/ɪ/-/ɛ/, /ɛ/-/æ/) and high-mid vowel and mid-low continua (/ʊ/-/ʌ/, /ʌ/-/ɒ/). In each continuum, the first three formants of the eight in-between sounds varied in each step with duration and F4 held constant. Each tenfold continuum is to be synthesized by PRAAT from one sound (SM, start mother sound) to the other (EM, end mother sound). Every continuum is numbered 1 to 10 from SM to eight in-betweens and then to EM.

Instead of adding repetitions in the stimuli pool, as in Mora [14], this experiment synthesizes its eight in-between sounds by augmenting ±11.11% of the disparity between the two mother sounds every step from SM to EM, and EM to SM sounds. According to the numbering rule of synthesized sounds of each continua, the two directions of one pair of continua should be in reverse order. Take /ɛ/-/æ/ for example, one continuum is from SM /ɛ/ to EM /æ/, numbered 1 to 10 respectively, and the other from SM /æ/ to EM /ɛ/, also numbered 1 to 10.

In the identification experiment, three duplicates of SM and EM sounds were added in each continuum as filters, with the purpose to exclude those responses done carelessly. A total of 104 stimuli were presented for forced-choice in the identification task (4 continua x (10 stimuli x 2 directions + 3 duplicates x 2 mother sounds)).

Following AXB discrimination of Mora [14] AXB task was designed with a two-step resolution and four orders of AXB, namely AAB, ABB, BBA, BAA. A total of 256 triplets of stimuli were created (8 triplets x 4 orders x 4 continua x 2 directions; ISI=1sec).

2.3. Design and Procedure

A Matlab based program was used to conduct the two experiments in a quiet room in a session lasting about 45 minutes. Each subject would receive a brief training to guarantee that he/she had fully understood the instructions before starting the test. The subject was then required to accomplish the identification task and AXB discrimination task. During the whole experiment, subjects heard only once the stimuli over headphones. They were encouraged to select their responses as fast as possible. Stimuli in both tasks were presented in fully randomized blocks.

3. RESULTS AND DISCUSSION

3.1. Identification Task

In this experiment where a total of 104 stimuli were judged by each subject whether it is closer to SM or EM sound. Correctness of filter vowels was used to exclude those whose score is below 85%, so that reliability of choice can be guaranteed. In our study, correctness of filters of all subjects is 0.88 (S.D.=0.072), guaranteeing that each subject paid due attention to the test during the whole process.

Theoretically, of the ten sounds in each continua, numbers 1 to 5 would be responded as "SM", and numbers 6 to 10 as "EM". The average of numbers which are not realized as expectations are defined as blur area (BA), which differs from "50% crossover" for the boundaries in [14] in that subjects in our experiments tend to have an area wider than 50% crossover. According to the numbering symmetry in our stimuli, BAs of the given contrast in two directions should total 11. Table 2 shows the BAs and Sum (S) of two BAs of the 4 continuum x 2 directions.

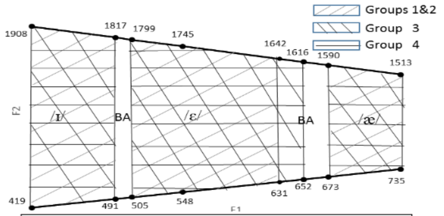
Table 2: BAs of 8 Vowel Continua; "S" refers to the sum of BAs of the two directions.

Group	Front Vowel						Back Vowel					
	/ɪ/→ /ɛ/	/ɛ/→ /ɪ/	S	/ɛ/→ /æ/	/æ/→ /ɛ/	S	/ʌ/→ /ʊ/	/ʊ/→ /ʌ/	S	/ʌ/→ /ɒ/	/ɒ/→ /ʌ/	S
1	6.5	4.5	11.0	6.0	5.0	11.0	5.5	6.0	11.5	5.5	6.5	12.0
2	6.5	4.5	11.0	6.0	5.0	11.0	6.5	6.0	12.5	5.5	6.5	12.0
3	6.5	4.5	11.0	6.5	5.0	11.5	6.5	5.5	12.0	5.5	5.5	11.0
4	6.5	4.5	11.0	6.5	5.0	11.5	6.5	4.5	11.0	4.5	6.5	11.0

The results conform to the hypothesis that BAs of the same vowel contrast, when synthesized in two opposite directions, manifest symmetrical characteristics, though variations could be found in pairs whose total is bigger than 11.

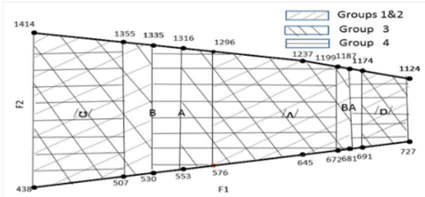
BAs of the four pairs of continua lingered around Steps 5 and 6, which help us to find out the range within which the three target sounds in /ɪ/-/ɛ/-/æ/ and in /ʊ/-/ʌ/-/ɒ/ can still be perceived as the original sound, given changes made in the first three formants. Such perception ranges can be projected in a continuum, shown in Figures 1 and 2, where the upper graphs show the boundaries of the first two formants (F1 and F2), and the lower tables presents the ranges (in percentage) done to the original target vowels when they can still be identified as the original, though a certain percentage of changes have been done to it. The percentage range is based on changes done to F1, since F2 changes accordingly in our experiment for a given F1.

Figure 1: Ranges of Perception for Front Vowels (in Hz and in %)



	/ɪ/	/ɛ/	/æ/
1&2	0~17.18%	-7.85~15.15%	-8.44%~0
3	0~17.18%	-7.85~19.22%	-8.44%~0
4	0~17.18%	-7.85~19.22%	-8.44%~0

Figure 2: Range of Perception for Back Vowels (in Hz and in %)



	/ʊ/	/ʌ/	/ɒ/
1&2	0~21.00%	-10.70~5.58%	-4.95%~0
3	0~21.00%	-14.26~5.58%	-4.95%~0
4	0~15.75%	-17.83~4.17%	-6.33%~0

The shades in Figs 1 and 2 suggest that, as language proficiency increases, the perception range gets wider for mid-vowel /ɛ/ and /ʌ/, from a range of 24% (15.15%-(-7.85%)) for beginners and intermediate level, to 27.07% (19.22%-(-7.85%)) for advanced level and native speakers for / ɛ /, and from 16.28% (5.59%-(-10.70%)) for beginners and intermediate level, to 19.84% (5.58%-(-14.26%)) for advanced level, and to 22% (4.17%-(17.83%)) for native speakers of English for /ʌ/, while changes are comparatively smaller in the ranges for high and low vowels. In the following section, such a hypothesis will further be discussed that the perception of L2 can be positively related to language proficiency.

3.2. Discrimination Task

In AXB task, 256 triplets were played for identification. Subjects were required to judge which sound of each triplet is different from the other two. Correctness of AXB discrimination task is analyzed from the aspects of spectral distance, cross-language vowel systems, and language proficiency levels.

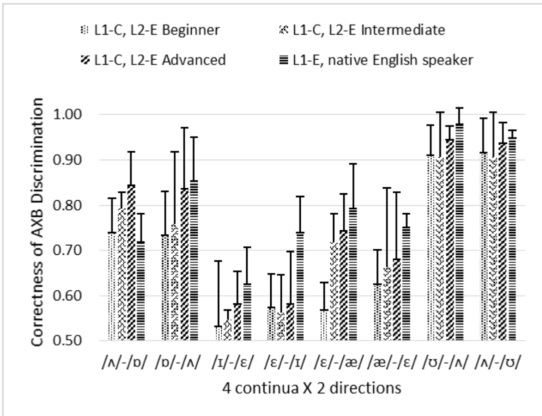
3.2.1. Spectral Distance

Distribution of the six mother sounds, i.e. /ɪ/, /ɛ/, /æ/, /ʊ/, /ʌ/, /ɒ/, results in different spectral distances of the four continua. Distance (SpD) between two vowels is calculated by

$$SpD = \sqrt{(F1_2 - F1_1)^2 + (F2_2 - F2_1)^2}. \quad (1)$$

The four contrasts follow the ranking of SpD's in the following order: SpDʌ-ɒ<SpDɪ-ɛ<SpDɛ-æ<SpDʊ-ʌ (139, 207, 299, 410, respectively, based on the formant values in [10]). Correctness of AXB discrimination didn't show a discernible difference between high and low spectral distance continua (Fig. 3).Such findings differ from [8] which stated that the more spectrally different the two vowels were, the easier they were to be discriminated.

Figure 3: Correctness of AXB Discrimination



The tendency in Fig.3 indicates that Chinese EFL learners perceived front and back vowel contrasts in a noticeably different way, with mid-low and mid-high back vowel contrasts being perceived less categorically than mid-low and high-mid front vowel contrasts. Such results partly conform to the statements from previous research that spectral disparity of two vowels is one factor contributing to the correctness of vowel perception [4, 8].

3.2.2. Cross-language Vowel Systems

Correctness of AXB by Chinese subjects is generally lower than that of native speakers, as shown in Fig. 3. In particular, bigger differences can be found in front vowel than in back vowel perception among the four groups. For example, correctness for /ɛ/→/æ/ pair is 0.53, 0.55, 0.58, and 0.63, respectively, as compared to 0.73, 0.76, 0.84, and 0.85 for /ɒ/→/ʌ/ pair.

Cross-language phonological realization between L1 and L2, i.e. differences of the two vowel systems, can explain this tendency. Mandarin Chinese (MC) includes six vowels, namely, [a], [o], [ɤ], [i], [u], and [y] [12], with [a] having either 13 phonological realizations [12], four [11] or five variants [5], according to different scholar' criteria. A consensus can be reached that /ɛ/ and /æ/ are two allophones of [a] in Chinese [5, 12]. According to Lin [12], /ɛ/ in [iɛn] '*yán*, salt' and /æ/ in [yæn] '*yuán*, round' are the phonological realizations of /ɛ/-/æ/ in Chinese. These two sounds are independent phonemes in English, while two allophones of one phoneme [a] in MC. Such a situation may well be explained by PAM-L2 [2] that Chinese EFL learners either fail to establish a new perceptive category for alien sounds, or assimilate perceptive categories with native sound system. Therefore, they tend to have lower correctness (average: 0.55) when distinguishing /ɛ/-/æ/ contrast.

Of the other four English phonemes, /ɪ/, /ʌ/, /ɒ/, /ʊ/, three can find their similar counterparts in MC ([i]-/ɪ/, [o]- /ɒ/, [u]- /ʊ/), with no /ʌ/-like sound in MC [12]. Results in Fig. 3 show that subjects perform better in continua containing alien vowel /ʌ/ than in continua solely containing either similar vowels (/ɪ/-/ɛ/) or same vowels realized as allophones (/ɛ/-/æ/). Such a trend finds explanation in PAM that L2 learners establish new perceptive categories for alien sounds [1], an explanation which echoes what SLM states that new sounds are easier than similar sounds to distinguish [8]. The impact of cross-language vowel systems on vowel perception is not only proved by subjects speaking two languages from identical language family as [1, 2, 8] did, but also by subjects speaking two languages from fraternal language families, i.e. MC and English.

3.2.3. Language Proficiency

A rough tendency can be noticed in Section 3.1 that as language proficiency increases, the perception range gets wider. This section is to further test whether language proficiency is a potential impact factor of AXB discrimination in vowel perception.

AXB discrimination scores were submitted to one-way ANOVA of average correctness (4 continua: /ɪ/-/ɛ/, /ɛ/-/æ/, /ʊ/-/ʌ/, and /ʌ/-/ɒ/), with language proficiency (4 levels: L2-E Beginner, L2-E Intermediate, L2-E Advanced, and L1-E native English speaker) as between-subjects factors. The main effect of contrast did not reach significance for all four continua ($F(3, 28)=0.780$ $p=0.52$). Such a factor was found to be moderately significant for front vowel contrasts /ɛ/-/æ/ and /ɪ/-/ɛ/ ($F(3, 12)=3.365$, $p<0.10$; $p=0.055$). Tukey's post hoc procedure indicates that 1) Group 4 recalled moderately significant higher correctness in front vowel contrasts than Group 3 (MD=0.08, $p<0.10$), Group 2 (MD=0.10, $p<0.10$), and Group 1 (MD=0.15, $p<0.05$), and 2) a moderately significant difference between Groups 2 and 3 (MD=0.02, $p<0.10$). However, significance failed to be reached for back vowel contrasts ($F(3, 12)=0.189$, $p=0.90$).

Such findings indicated that performance of subjects with different language proficiency varied in the perception of high-mid and mid-low front vowel contrasts, that is, as language proficiency is enhanced, front vowels are perceived more categorically, while back vowel contrasts didn't show such a tendency.

4. CONCLUSION

This study, by comparing the perception of English /ɪ/-/ɛ/, /ɛ/-/æ/, /ʊ/-/ʌ/, and /ʌ/-/ɒ/ contrasts by 16 Chinese EFL learners of varied English proficiency, and by four native English speakers, support the hypothesis that spectral distance, cross-language vowel system and language proficiency are three cues affecting vowel production for Chinese EFL learners.

Results reveal that the disparity in the correctness of front and back English vowel contrasts for Chinese EFL learners is saliently higher than that for native English speakers, such a finding can be explained by the differences in the phonological realizations of the same stimuli in English and MC vowel systems. Furthermore, positive correlation shows itself only when stimuli are compared separately according to their places of articulation (front or back), that is, in case of either front vowels or back vowels, contrasts with bigger spectral distance are perceived more accurately than those with smaller spectral distance. Such a phenomenon proves that perceptive categories of similar sounds between two sound systems are apt to be mingled while alien sounds establish new perceptive categories. Finally, language proficiency is proved to be related to vowel perception for two clues: 1) perception range is proved to be wider as language proficiency increases, especially for mid-vowels; 2) correctness of front vowel perception is proved to have a moderately positive correlation with language proficiency. We have to admit that with a larger data pool, the results would be more convincing in the analysis of the relation between language proficiency and vowel perception.

5. Acknowledgements

This work is supported in part by the National Basic Research Program of China (No. 2013CB329301), and in part by the National Science Foundation of China (No. 61233009), and in part by the Social Science Foundation of Tianjin, China (No. TJWW13-001).

6. REFERENCES

[1] Best, C. 1995. A direct realist view of cross-language speech perception. In: Strange W. (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press: 171–204.

[2] Best, C., Tyler, M. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: O.-S. Bohn, M. Munro (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege*. 13–34. Amsterdam: John Benjamins.

[3] Bohn, O.-S., Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11, 303–328.

[4] Brasileiro, I. 2007. Language mode and bilingual's speech perception. In C. Keskin, M. Kolkman, M. Rigterink (Eds.), *UiL OTS yearbook 2006*. Utrecht: LOT:19–33.

[5] Duanmu, S. 2007. *The phonology of standard Chinese*. Oxford: Oxford University Press.

[6] Escudero, P., Benders, T., Lipski, S. C. 2009. Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics* 37(4), 452-465

[7] Escudero, P., Simon, E., Mitterer, H. 2012. The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics* 40(2), 280-288.

[8] Flege, J. E., Bohn, O. S., Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics* 25(4), 437-470.

[9] Ho, Y. K. 2010. Technical university freshmen's problems in English segmental pronunciation—a case study of English front vowel perception. *Minghsin Journal*, 36(2), 33-44.

[10] Ladefoged, P., Johnson, K. 2014. *A course in phonetics*. Cengage learning, 44.

[11] Lin, C. Y. 2014. Perception and Production of Five English Front Vowels by College Students. *English Language Teaching* 7(9), 14-20.

[12] Lin, T., Wang, L. J. 2013. *A Course of Phonetics*. Beiijng: Peking University Press, 233-239

[13] Mora, J. C., Fullana, N. 2007. Production and Perception of English /i/-/ɪ/ and /æ/-/ʌ/ in a Formal Setting: Investigating the Effects of Experience and Starting Age. *Proceedings of the 16th international congress of phonetic sciences* 3, 1613-1616.

[14] Mora, J. C., Keidel, J. L., Flege, J. E. 2010. Why are the Catalan contrasts between /e/-/ɛ/ and /o/-/ɔ/ so difficult for even early Spanish-Catalan bilinguals to perceive. *Achievements and perspectives in the acquisition of second language speech: New Sounds* 2, 183-193.

[15] Mora, J. C., Nadeu, M. 2012. L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism* 16(4), 484-500.

[16] Tseng, W. C. 2011. A study of Taiwanese high school students' production and perception performance in English non-high front vowels. Retrieved from http://ndltd.ncl.edu.tw/cgibin/gs32/gsweb.cgi/login?o=dnclcdr&s=id=%22099STUT8741022%22.&searchmode=basic

[17] Wang, H., van Heuven, V. J. 2004. Cross-linguistic confusion of vowels produced and perceived by Chinese, Dutch and American speakers of English. *Linguistics in the Netherlands*, 21(1), 205-216.