Temporal and Formant Trajectory Analysis of English Tense-Lax Vowels Produced by Native Chinese Speakers

Jian Gong*,§, Di Xue†, William Bellamy*, Feng Wang* and Xiaoli Ji*
*School of Foreign Languages, Jiangsu University of Science and Technology, Zhenjiang, China
†SIP Xingwan School, Suzhou, China
§ E-mail: j.gong@just.edu.cn Tel/Fax: +86-0511-84401945

***Abstract*—The current study investigated the acoustic properties of English tense-lax vowels produced by native Chinese speakers. A series of acoustic parameters such as vowel duration, formant frequencies, and the dynamic formant trajectories were measured and analyzed. The results showed that Chinese speakers could produce English tense-lax vowels with significant duration differences. However, no significant spectral differences were found between the two types of vowels. Further analysis illustrated strong L1 influence for the Chinese speakers on duration, formant frequencies, and formant trajectories. These results indicated that the Chinese speakers largely ignored the important spectral quality differences between English tense and lax vowels. They relied on the duration feature in distinguishing the English tense-lax contrasts.**

I. Introduction

Vowel acquisition is an important aspect in second language (L2) sound acquisition. Different from consonants, the transitions between vowels are continuous. Usually, we use tongue position and lip protrusion (e.g., front-back, high-low, rounded-unrounded) to describe different vowels. However, there is no clear boundary between the adjacent vowels. For learners, learning an L2 that has more vowel inventories than their native language could cause great trouble [1], [2]. For example, Mandarin Chinese has 6 monophthongs [3], while the number of monophthongs in English is 12 [4]. Therefore, for Chinese learners of English, discriminating English vowels in a relatively crowded vowel space could be a big challenge.

Some crucial acoustic features are the most important cues in distinguishing different phonetic/phonology categories. These features are normally the keys for successful L2 sound acquisition [5]. For example, Chinese speakers have problems differentiating English voiced/voiceless fricatives, because they have not acquired the voiced/voiceless feature in English [6], [7]. Similarly, Spanish speakers have problems correctly producing English voiced/voiceless plosive contrasts because the VOTs are quite different in the two languages [8]. For vowels, this kind of distinctive feature normally not only exists in one but in multiple dimensions. For example, English tense and lax vowels not only have differences in spectral quality but also in duration.

First language (L1) influence is possibly one of the most important single factors in successful L2 sound acquisition. An important aspect of L1 influence is that, for the learners, some crucial distinctive features in L2 do not exist in L1, or are in L1 but used differently. Therefore, the learner may ignore or misuse these features, and the difficulty in perception and production of some L2 sounds can be expected [9]. Classic theoretical models in L2 sound acquisition such as PAM [10], SLM [11] and NLM [12] all claim that learners process L2 sounds via their native categories. The most difficult sounds in an L2 are those sounds which have close counterparts in L1 but with subtle differences, because these subtle differences may be trivial in L1 but crucial in L2.

In most previous vowel studies, static acoustic measurements such as duration and formant values (either mean or midpoint values) were used. However, some studies found that even for monophthongs the formant values were changing during vowel production [13]. This spectral change is context-independent and is a systematic property of the vowel itself. More recently, researchers have started to apply the formant dynamic analysis in the studies of dialectal variations of American English vowels [14] and the development of L2 vowel production by bilingual Children [15]. The current study aimed to investigate native Chinese speakers' production of English tense and lax vowels. Both static and dynamic measures were applied in the current study and the focus were set on L1 influence.

II. Methods

A. Speakers

Fifty native Chinese speakers (31 females and 19 males, mean age = 24 years, age range: 21-28 years) were recruited from Jiangsu University of Science and Technology, China. These speakers were all originally from the *Jianghuai* Mandarin dialect-speaking region and none had reported any language or hearing problems. Speakers were paid for their participation. According to their College English Test-Band 6 (CET-6) scores, these speakers were divided into two groups: a low experience group (LE) and a high experience group (HE). The CET-6 is a national, large-scale standardized test, supervised by the Ministry of Education of China. As a criterion-related norm-referenced test, the CET-6 aims to assess students English proficiency. The LE group comprised 25 speakers with a mean CET-6 score of 419 ($SD = 24$). The HE group consisted of 25 speakers whose mean score on the CET-6 was 503 ($SD = 36$). The two groups differed significantly in terms of their CET-6 scores ($p < 0.001$). Three native British male speakers from south England who worked in China also participated in the current study.

B. Materials

For the English recording materials, 4 tense vowels (/ɑ: i: ɔ: u:/) and 5 lax vowels (/æ ʌ ɪ ɒ ʊ/) were used in the current study. Real English words with the 9 vowels in /bVd/ frame (*bad, bard, board, bead, booed, bud, bid, bod, buddha*) were recorded (with *buddha* in a /bVda/ frame as an exception). For the Chinese materials, the focus was set on the 4 vowels (/a i o u/) which had relatively close counterparts in English. Four nonsense Chinese phrases each consisting of two characters (八的, 逼的, 播的, 不的) were chosen as recording materials. The pronunciation of the first characters in these phrases are /ba, bi, bo, bu/ respectively, while the pronunciation of the second character is /də/. In this way, we simulate the English /bVd/ structure in Chinese.

C. Procedure

The recordings were conducted in a sound-treated laboratory in Jiangsu University of Science and Technology. Chinese speakers first finished the Chinese vowel recordings and then the English one. The recordings were controlled by an experimenter using a customized MATLAB program. During the recording period, English /bVd/ words or Chinese phrases were shown on a computer screen one by one in a random order, and the speakers were asked to read these words or phrases aloud using a normal voice and speed. An AKG C4000B microphone and an RME Fireface 800 sound card were used in the recordings, with a WAV recording format, a sampling rate of 44.1kHz, and the quantization depth was set at 16bit. Similar procedures were followed by the native English speakers for the 9 /bVd/ words.

D. Acoustic measurements

Praat [16] was used to annotate all the recorded tokens as well as to extract the acoustic parameters (i.e., duration and formant values). F1 and F2 values were measured at five equidistant temporal locations (20-35-50-65-80% point) of the vowel part for each token, to capture the dynamic spectral change of the vowel movement. To eliminate variation caused by physiological differences among speakers, a Bark Difference Metric method [17] was applied to normalize the raw formant frequency values by converting them into the Bark scale using the formula:

$$Z_i = 26.81/(1 + 1960/F_i) - 0.53 \tag{1}$$

Where F_i is the value for a given formant i in Hz, and Z_i is the Bark-converted value. The difference of Z3 - Z2 is used to represent the normalized front-back dimension and Z3 - Z1 is used to represent the normalized height dimension.

To measure the dynamic spectral change, trajectory length (TL) [15] values were calculated for each vowel tokens. In the current study, the trajectory length was computed by using the normalized F1 (Z3-Z1) and F2 (Z3-Z2) values, following the formula introduced in [15]:

$$TL = \sum_{n=1}^{4} VSL_n \tag{2}$$

where TL is the sum of the Euclidean distances (in normalized F1-F2 plane) between each two consecutive temporal points, (i.e., 20-35%, 35-50%, 50-65%, 65-80%), and the length of each vowel section (VSL) is calculated based on the formula:

$$VSL_n = \sqrt{(F1_{n+1} - F1_n)^2 + (F2_{n+1} - F2_n)^2} \tag{3}$$

III. Results and Analysis

A. Temporal analysis

The mean duration of English tense-lax vowels produced by native English speakers and the two groups of Chinese learners are shown in the upper panel of Fig. 1. For native speakers, the difference between tense-lax vowels was clear, that is, tense vowels have a longer duration than its lax vowel counterpart. For example, the vowel durations for /ɑ:-ʌ/, /i:-ɪ/, /ɔ:-ɒ/, /u:-ʊ/ were 322-155, 269-141, 311-166, 272-148 (ms) respectively. Similar to native speakers, clear differences between tense-lax vowels were also found for Chinese learners. Two-sample t-tests confirmed that there were significant differences between /ɑ:-ʌ/, /i:-ɪ/, /ɔ:-ɒ/, /u:-ʊ/ for the HE (High Experience) group, while for the LE (Low Experience) group, except for a marginally significant difference between /ɔ:-ɒ/ ($p = 0.065$), significant differences were found for the remaining 3 pairs of vowel contrasts ($p < 0.05$).

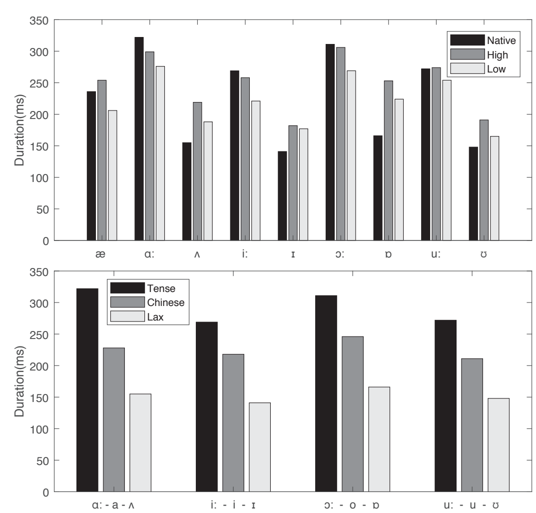

Fig. 1. Durations of tense-lax vowels produced by native English speakers and the two groups (high/low experience) of Chinese speakers (upper panel). Note that the Chinese vowel durations are shown along side native English vowel durations in the lower panel.

Another interesting result that can be observed from Fig. 1 is that the duration of tense vowels for the HE group was closer to native speakers than the LE group, while the durations of lax vowels produced by the LE group were closer to native speakers than that of the HE group's. T-test analysis demonstrated that, for the HE group, there was no significant difference of tense vowel duration (/æ ɑ: i: ɔ: u:/) between them and the native speakers ($p > 0.05$). However, their lax vowels (/ʌ ɪ ɒ ʊ/) were significantly longer than the native speakers ($p < 0.05$). For the LE group, statistical analysis revealed that only three vowels were not significantly different from native speakers (/ɑ:/, $p = 0.06$; /u:/, $p = 0.35$, /ʊ/, $p = 0.07$).

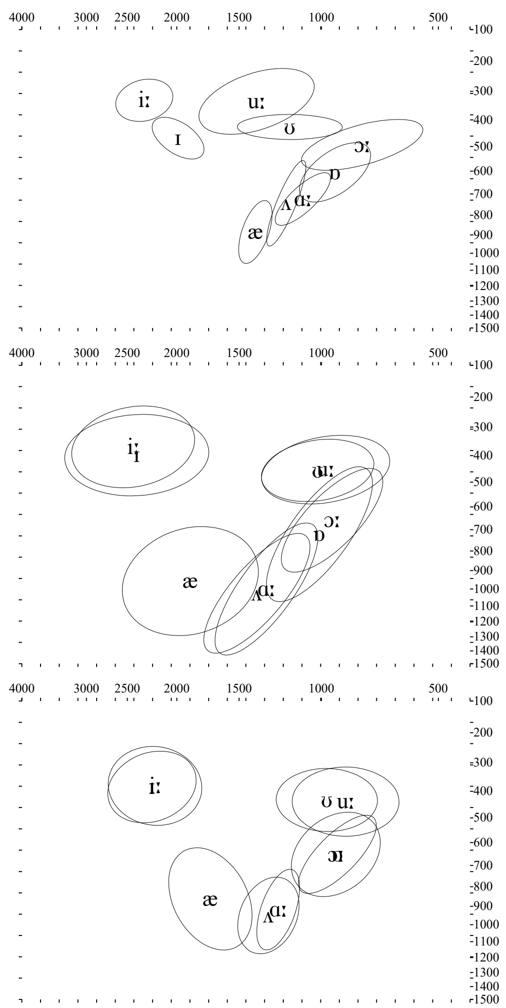

Fig. 2. Overall distribution of English tense-lax vowels produced by native English speakers (upper), HE group (middle) and LE group (lower) in the F1-F2 plane.

To investigate the possible L1 influence, mean durations of the 4 Chinese vowels (/a i o u/) across all speakers in both HE and LE groups are shown in the lower panel of Fig. 1, alongside the duration of their English tense and lax counterparts from the native speakers. It can be seen that the duration of Chinese vowels was in-between their English tense-lax counterparts, i.e., shorter than the tense vowels but longer than the lax vowels. This result suggests that the two groups of Chinese learners' production of English tense-lax vowels were influenced by their native language. In most of the cases, especially for the LE group, Chinese learners' tense vowels were shorter than natives' while their lax vowels were longer than natives'.

To summarize, the temporal analysis revealed that Chinese learners could differentiate English tense-lax vowel contrasts from the duration dimension, but the clear L1 influence was also evident.

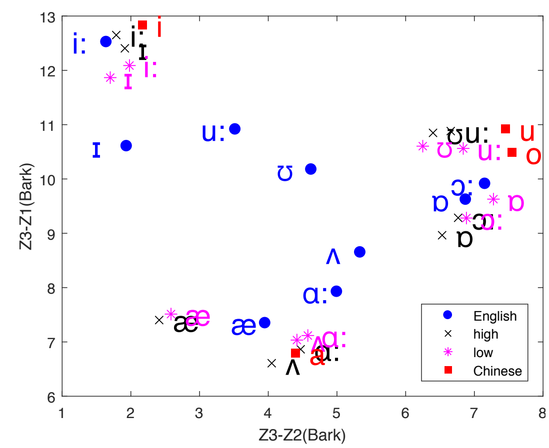

Fig. 3. Mid-point formant values for English tense-lax vowels from native speakers and Chinese speakers (both high/low experience groups). Note that the formant values for the 4 Chinese vowels are also given.

B. Overall vowel space

Fig. 2 shows the overall distribution of the 9 English tense-lax vowels in the F1-F2 space for the three groups of speakers. The ellipses were drawn using Praat based on the midpoint F1-F2 values from each vowel across all speakers. The axes were labeled in Hz but scaled in Bark. From the upper panel of Fig. 2, it can be seen that although there were some overlaps, the boundaries between tense and lax vowels were relatively clear for native English speakers, and each vowel did not span a large distribution area but was quite concentrated. In contrast with the native speakers, both HE (middle panel) and LE (lower panel) groups of Chinese speakers showed larger distribution area for each vowel, and their tense vowels were highly overlapped with their lax counterparts. The distribution patterns between the HE and LE speakers were quite similar, though the HE group showed a slightly larger distribution area than the LE group. This result indicates that both groups of Chinese speakers did not differentiate the English tense-lax vowel contrasts and treated them almost the same in production, at least from the spectral aspect.

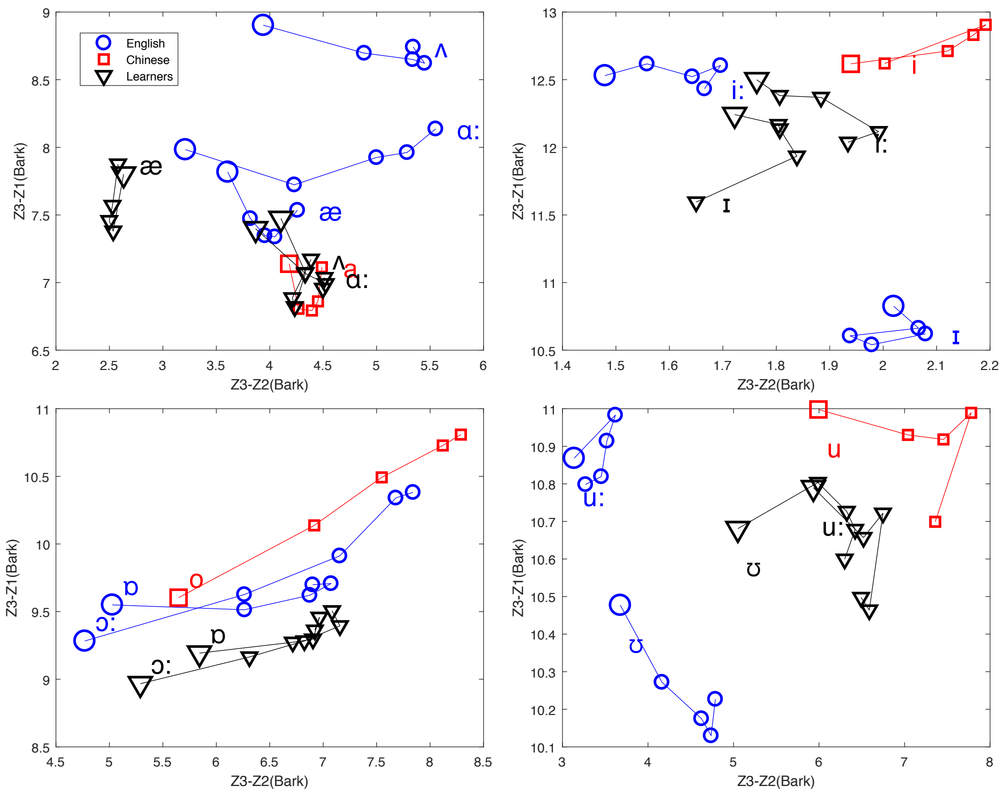

Fig. 4. Formant trajectories for English tense-lax vowels from native English speakers and Chinese speakers (mean across high/low experience groups). Note that the formant trajectories for the 4 Chinese vowels are also given.

C. Midpoint formant distributions

To further compare Chinese speakers' English vowels production with native English speakers', midpoint formant distributions of the 9 English tense-lax vowels from all three groups are shown in Fig. 3. Meanwhile, to investigate the L1 influence, midpoint formant distributions of the 4 Chinese vowels are also shown in Fig. 3. Again from Fig. 3 we can see that there are relatively large distances between most of the tense and lax vowel pairs (/ɑː-ʌ/, /iː-ɪ/, /uː-ʊ/) produced by native English speakers. Among the 4 Chinese vowels, /i/ was quite close to English /iː/, and /o/ was close to English /ɔː/. Chinese /u/ was located relatively further back than the English /uː-ʊ/ contrast while Chinese /a/ was lower than its English counterparts (/ɑː, æ, ʌ/). This distribution pattern was consistent with many previous studies [14].

Consistent with Fig. 2, Fig. 3 demonstrates that the Chinese speakers of both HE and LE groups produced quite similar tense-lax /ɑː-ʌ/, /iː-ɪ/, /ɔː-ɒ/, /uː-ʊ/ contrasts. T-tests demonstrate that there was no significant difference in either front-back (Z3-Z2) or high-low (Z3-Z1) dimensions for these contrasts ($p > 0.05$). Clear L1 influence can also be observed from Fig. 2 in that the English /ɑː-ʌ/ contrast produced by Chinese learners was close to Chinese /a/, while English /iː-ɪ/ and /uː-ʊ/ contrasts were close to Chinese /i/ and /u/ respectively. Again, this result suggests that Chinese learners didn't differentiate the English tense-lax contrasts from the spectral aspect, and relied on their native categories to produce these sounds.

D. Formant trajectories

Fig. 4 shows the formant trajectories for the English and Chinese vowels from native English speakers and the two groups of Chinese speakers. The trajectories of English vowels produced by Chinese speakers were calculated using the mean normalized formant values across the HE and LE groups. From the upper left panel of Fig. 4, it can be seen that the trajectory patterns for native English /æ, ɑː, ʌ/ were quite different, with /æ/ a bit similar to Chinese /a/. Chinese speakers' English /ɑː, ʌ/ patterns were similar to each other but different from both Chinese /a/ and native English /æ, ɑː, ʌ/. In the upper right panel, although the trajectory patterns for English /iː, ɪ/ produced by Chinese speakers were similar to each other, they were not similar to either naive English /iː/ nor Chinese /i/, and clear differences can be observed from native English /ɪ/. Similar situations can be found in the lower left and lower right panels of Fig. 4, that is, the trajectory patterns of English tense-lax vowels produced by Chinese speakers were similar to each other, and maybe also similar to Chinese vowels or native English tense vowels, but different from the native English lax vowels. Again, this result indicates strong L1 influence in the production of English tense-lax vowels for Chinese speakers.

E. Trajectory length

Fig. 5 shows the mean trajectory length for the English tense-lax vowels and Chinese vowels from native English and Chinese speakers. Again, large trajectory length differences between native English speakers and Chinese speakers can be observed. Statistical analysis confirmed that there were significant differences for almost all the English tense-lax vowels (/ɑː, ʌ, iː, ɪ, ɔː, ɒ, uː, ʊ/, all $p < 0.05$) between the two groups of speakers. Fig. 5 also shows that the Trajectory length of English /ɑː/ and /iː/ produced by Chinese speakers was close to the Chinese /a/ and /i/ respectively. T-tests confirmed that there was no significant trajectory length difference between these sounds, indicating the possible influence of L1.

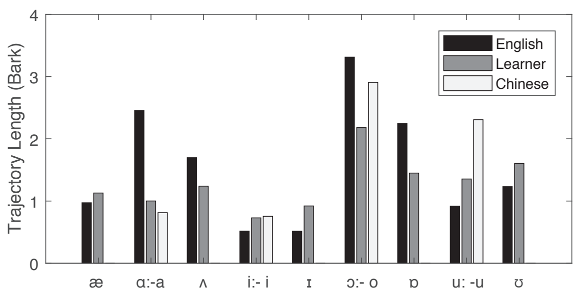

Fig. 5. Trajectory lengths for English tense-lax vowels from native English speakers and Chinese speakers (mean across high/low experience groups). Note that the trajectory lengths for the 4 Chinese vowels are also given.

IV. Discussion

The current study investigated the acoustic properties of English tense-lax vowels produced by native Chinese speakers. A series of acoustic parameters such as vowel duration, formant frequencies, and formant trajectories were measured and analyzed. Significant differences of vowel duration were found between English tense-lax vowel pairs, indicating that Chinese speakers did discriminate English tense-lax contrasts in the duration dimension. However, the results also showed large duration differences between Chinese speakers (especially for the LE group) and native English speakers for most of the vowels, which is to say, they tended to have a shorter duration for tense vowels but longer duration for lax vowels. Further analysis revealed that the duration of Chinese vowels was in between their English tense-lax counterparts. These results indicate that although Chinese speakers were sensitive to the duration and had the awareness of the duration differences between English tense-lax vowels, their production was still strongly influenced by their native language. The L1 categories had some possible “magnetic” effects on Chinese speakers’ L2 categories as introduced in NLM [12].

Spectral analysis results revealed that Chinese speakers’ production of English tense-lax vowel contrasts were highly overlapped with each other in the F1-F2 plane. As suggested by PAM [10], if listeners assimilated two non-native sounds into one single native category, great difficulty in discriminating the two non-native sounds could be predicted. Further analysis of the results confirmed that the Chinese speakers assimilated the English tense-lax vowel contrast into one Chinese sound, and therefore treated them as the same sound, and used almost the same method to produce them. In SLM [11], Flege claims that only if the subtle difference between the L2 category and the similar L1 category can be perceived by the learner, then a new category for the L2 sound can possibly be established by the learner, and its production becomes correct. In the current study, it is clear that the Chinese speakers did not appreciate the differences between English tense-lax vowels, so they referred to the closest Chinese category in producing these sounds.

The formant trajectory analysis in the current study investigated L2 production from a different perspective. The results were consistent with the findings from the static formant and temporal analyses. Chinese speakers demonstrated quite similar spectral dynamics in English tense-lax vowel production, and some evidence of L1 influences were also found. However, the formant trajectory patterns in the current study were not as smooth as those reported in some previous studies [14], [15]. This might be due to the limited sample size and individual differences in the current study, especially for the native English speakers. Future work should employ larger speaker groups and the dialect background of the speakers should be controlled more tightly.

In general, the current study demonstrated the L1 (Mandarin Chinese) influence on speakers’ L2 English vowel production. Mandarin has many fewer monophthongs than English, and therefore it is difficult for Chinese speakers to discriminate the sounds in the more crowded English vowel space. More specifically, in English, there are many tense-lax vowel contrasts, whilst tense-lax is not a distinguishing feature in Mandarin. Chinese speakers tend to rely on other features they are more familiar with to distinguish the English tense-lax contrasts, in this case, duration. However, the duration is only a secondary distinguishing feature for English tense-lax contrasts, and sometimes it is not stable and reliable [18]. This causes great difficulty for Chinese speakers in discriminating tasks [19] and also leads to Chinese speakers producing English tense-lax contrasts with clear duration differences but with almost the same spectral quality, as found in the current study.

Acknowledgment

This study was supported by grants from the Humanities and Social Science Fund of the Ministry of Education of China (16YJC740020, 15YJC740034), the Jiangsu Social Science Fund (18YYB009, 19YYC001).

References

[1] R. L. Bundgaard-Nielsen, C. T. Best, and M. D. Tyler, “Vocabulary size is associated with second-language vowel perception performance in adult learners,” *Studies in Second Language Acquisition*, pp. 433–461, 2011.

[2] P. Iverson and B. G. Evans, "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers," *Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 866–877, 2009.
[3] W. Lee and E. Zee, "Standard Chinese (Beijing)," *Journal of the International Phonetic Association*, vol. 33, no. 1, pp. 109–112, 2003.
[4] P. Roach, "British English: Received Pronunciation," *Journal of the International Phonetic Association*, vol. 34, no. 2, pp. 239–245, 2004.
[5] J. Gong, M. L. G. Lecumberri, and M. Cooke, "Ab initio perceptual learning of foreign language sounds: Spanish consonant acquisition by chinese learners," *System*, vol. 66, pp. 142–155, 2017.
[6] J. Gong, M. Cooke, and M. L. G. Lecumberri, "A quantitative model of first language influence in second language consonant learning," *Speech Communication*, vol. 69, pp. 17–30, 2015.
[7] J. Gong and W. Zhou, "Effect of experience on Chinese assimilation and identification of English consonants.," in *Proc. ICPhS XVIII*, 2015.
[8] J. E. Flege, "Age of learning affects the authenticity of voice-onset time (vot) in stop consonants produced in a second language," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 395–411, 1991.
[9] J. Gong, Z. Yang, X. Ji, and F. Wang, "Investigating the effectiveness of auditory training on Chinese listeners' perception of English vowels," in *Proc. ICPhS XIX*, pp. 2253–2257, 2019.
[10] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Theoretical and methodological issues* (W. Strange, ed.), pp. 171–204, Baltimore: York Press, 1995.
[11] J. E. Flege, "Second-language speech learning: Theory, findings, and problems," in *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (W. Strange, ed.), pp. 233–277, Baltimore: York Press, 1995.
[12] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson, "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (nlm-e)," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1493, pp. 979–1000, 2008.
[13] J. Harrington and S. Cassidy, "Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English," *Language and Speech*, vol. 37, no. 4, pp. 357–373, 1994.
[14] R. A. Fox and E. Jacewicz, "Cross-dialectal variation in formant dynamics of American English vowels," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2603–2618, 2009.
[15] J. Yang and R. A. Fox, "Acoustic properties of shared vowels in bilingual Mandarin-English children," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.3. 71)," *http://www. praat. org/*, 2014.
[17] E. R. Thomas and T. Kendall, "Norm: The vowel normalization and plotting suite," *[Online Resource: http://ncslaap.lib.ncsu.edu/tools/norm/]*, 2007.
[18] Y. Jia and Y. Wang, "Typology of English monophthongs by EFL learners from Wu dialectal region-A case study of Ningbo and Shanghai," in *Proc. 9th International Conference on Speech Prosody 2018*, pp. 843–847, 2018.
[19] D. Xue, J. Gong, W. Zhou, and X. Ji, "The effect of experience on Chinese listeners' assimilation and discrimination of L2 English vowels," in *2017 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Technique(O-COCOSDA)*, 2017.