Investigating the Utilizing of Durational and Spectral Cues in the Perception of English /ɪ-i:/ Contrast by Chinese Listeners

Yuhang Pei
School of Foreign Languages,
Jiangsu University of Science and Technology
Zhenjiang, China
pyh_kelia@163.com

Jian Gong *
School of Foreign Languages,
Jiangsu University of Science and Technology
Zhenjiang, China
j.gong@just.edu.com

***Abstract*—English vowel /ɪ/ and /i:/ are different in both durational and spectral dimensions. To investigate Chinese listeners' perceptual sensitivity to durational and spectral cues for the English /ɪ-i:/ contrast, and the effect from their L2 proficiency, two groups of native Chinese listeners with different English levels participated in the current study. A word recognition test with naturally produced /ɪ-i:/ tokens was used to examine listeners' English levels. Four categorical perception tests with synthesized /ɪ-i:/ continua were carried out to investigate listeners' ability of utilizing durational and spectral cues in /ɪ-i:/ perception. The results showed no clear categorical perception in the durational dimension for both groups, while in the spectral dimension, the high English proficiency group demonstrated a more categorical-like perception pattern. The results also showed that the high English proficiency group had better ability in utilizing spectral and durational cues for English /ɪ/ and /i:/ identification, indicating that listeners' cue-weighting strategies were influenced by their English levels.**

Keywords—Categorical Perception, Spectral and Durational Cues, Speech Perception

I. INTRODUCTION

Acquiring vowels in a second language (L2) is an important aspect for adult L2 learners. Unlike consonants, there is no clear boundary between the adjacent vowels. Therefore, many learners have great difficulty in identifying and discriminating L2 vowels [1]. Compared to other English vowel contrasts, English /ɪ/ and /i:/ are among the most difficult contrasts for learners [2],[3],[4].

Categorical Perception (CP) refers to the ability that listeners perceive continuous acoustic signals as discrete linguistic representations [5]. Related studies have been mainly documented on the categorical perception of consonants and vowels [6], [7]. Although the transitions between vowel categories are continuous, [7] showed that native listeners still exhibited categorical features in vowel perception.

Some acoustic features, such as spectral (frequencies of the first(F1) and the second(F2) formants) and durational cues, are the most important cues in distinguishing and identifying the phonetic category of /ɪ/ and /i:/. For native English listeners, they relied heavily on spectral cues with vowel duration at a secondary place [8]. However, some studies found that, unlike native English listeners, Japanese, Chinese and Spanish listeners relied predominantly on vowel duration to distinguish English /ɪ/ and /i:/ [8-10], while adult Korean listeners were able to employ both spectral and durational cues [11].

The cue-weighting strategies may conflict between the first language (L1) and the second language. Learners gradually use the strategies of cue weighting in L2 learning. [12] showed that L2 learning might change learners' cue-weighting strategies, specifically, L2 learners reduced their reliance on duration cues and improved their use of spectral cues in identifying English vowels.

Furthermore, L2 proficiency has exerted a crucial influence on categorical perception [13], [14]. For instance, a significant group difference was found between the high English proficiency group and the low English proficiency group. The high English proficiency group tended to make better use of acoustic cues to identify speech sounds. Specifically, Korean listeners of English in the high English proficiency group were more sensitive to F0 and Voice Onset Time (VOT) than listeners in the low English proficiency group [13],[14].

To summarize, previous studies have suggested acoustic cues of English vowel contrasts are important in vowel perception. However, there is a lack of studies on the relationship between acoustic cues and English proficiency, especially among Chinese listeners. Therefore, the goal of the current study is to investigate the perceptual ability and cue-weighting strategies of English vowel contrasts /ɪ/-/i:/ for Chinese listeners with different proficiency of English. More specifically, a word recognition test and four vowel identification tests were used in this study. Synthesized vowel continua based on original /ɪ/ and /i:/ in vowel identification tests were applied to examine the identification of English /ɪ/ and /i:/ among Chinese listeners with different English levels.

II. METHODS

A. Listeners

A cohort of twenty native Chinese listeners (14 females and 6 males) were recruited from Jiangsu University of Science and Technology, China, with age ranging from 22 to 28 years (mean age = 23.75 years). All listeners started learning English at school when they were 9 or 10 years old. They were all originally from the Northern Mandarin dialect-speaking region. None of them had reported history of speech or hearing disorders and a residence history in English-speaking countries. They all had passed the College English Test Band 4 (CET-4) and had basic knowledge of the International Phonetic Alphabet (IPA) for English. Listeners were paid for their participation. Following the procedure in

*Corresponding author

[15], these listeners were divided into a low English proficiency (LEP) group and a high English proficiency (HEP) group by using the k-means clustering algorithm, according to their word recognition scores (see section III A).

B. Stimuli

For the word recognition test, English tense-lax vowels (/ɪ/ - /i:/) were put in /CVC/ frame and 10 real words were used (*bid, bead, fit, feat, did, deed, rid, read, fid, feed*). Another 10 real words (*bet, bat, bed, bad, dead, dad, fed, fad, fet, fat*) were added as fillers. Six native British English speakers (2 females, 4 males) were instructed to produce these 20 words in a natural manner and to repeat them four times.

For the vowel identification tests, stimuli were synthesized based on the vowel parts (/ɪ/-/i:/) extracted from two real-word tokens (*hid, heed*), produced by another male native British English speaker. The duration values of the original /ɪ/ and /i:/ were 103ms and 178ms, and the first and second formant (F1, F2) values were 474Hz, 2089Hz and 343Hz, 2398Hz for /ɪ/ and /i:/ respectively. Vowel continua were constructed by using Vocal Toolkit for Praat [16] along durational and spectral dimensions with the original /ɪ/ as the starting point respectively. The first continuum was synthesized based on the original /ɪ/ where vowel duration was varied in 10 equal steps from 103ms to 178ms without changing the formants (/ɪ/ → /i:/ continuum). The second continuum was synthesized based on the original /i:/ with the duration reduced from 178ms to 103ms in 10 equal steps (/i:/ → /ɪ/ continuum). Similarly, another two continua (/ɪ/ → /i:/ and /i:/ → /ɪ/) were synthesized from the original /ɪ/ and /i:/ respectively, with F1 changing from 474Hz to 343Hz and F2 changing from 2089Hz to 2398Hz simultaneously in 10 equal steps, without modifying the original vowel duration. Finally, four sets of continua were constructed, each contained 11 stimuli (see sample stimuli in Fig. 1). The duration and formant values for stimuli in the four continua are shown in TABLE I. Continua based on another two vowels (/e/-/æ/) were created in a similar way and served as fillers.

TABLE I. DURATION AND F1 & F2 VALUES OF /ɪ/ - /I:/ IN THE CONTINUA

Step	F1(Hz)	F2(Hz)	Duration (ms)
1	474	2089	103
2	461	2119	111
3	448	2150	118
4	435	2181	126
5	422	2212	133
6	409	2243	141
7	395	2274	148
8	382	2305	156
9	369	2336	163
10	356	2367	171
11	343	2398	178

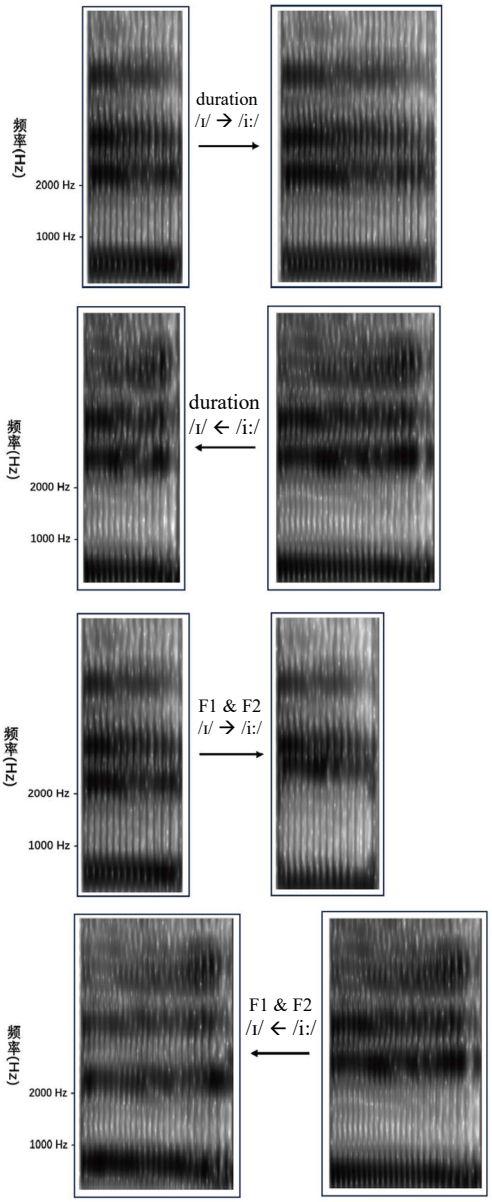


Fig. 1. Spectrograms of the endpoint stimuli in the four synthesized continua. The top panel shows the durational /ɪ/ → /i:/ continuum where the vowel length was increased from the original /ɪ/ to /i:/. The upper-middle panel shows the durational /i:/ → /ɪ/ continuum where the vowel length was decreased from the original /i:/ to /ɪ/. The lower-middle panel shows the spectral /ɪ/ → /i:/ continuum where the values of F1 and F2 were adjusted from the original /ɪ/ to /i:/. The bottom panel shows the spectral /i:/ → /ɪ/ continuum where the values of F1 and F2 were adjusted from the original /i:/ to /ɪ/.

C. Procedures

All tests were conducted in a sound-treated audiology test laboratory at Jiangsu University of Science and Technology. Listeners were tested one by one. The presentation of stimuli and the collection of responses were controlled by a tester using a customized E-prime [17] program. Listeners finished the English word recognition test first, followed by the vowel identification tests with the two durational continua, and then the vowel identification tests with the two spectral continua.

In the English word recognition test, listeners were asked to classify the English vowel they heard in each CVC token as an instance of one of the four vowel categories (/ɪ/, /i:/, /e/, /æ/) by clicking the responding button on the computer screen. Four English vowels (/ɪ/, /i:/, /e/, /æ/) were shown on buttons

to represent the corresponding vowel category. The word recognition test included 240 trials (10 words × 6 speakers × 4 tokens) and 240 fillers (10 English words × 6 speakers × 4 tokens) in a total of 480 stimuli. The order of the stimuli presentation was random.

In the vowel identification tests, listeners were required to assign the vowel they heard to their corresponding vowel category. The interstimulus interval (ISI) was 500ms. There were 1760 stimuli (20 repetitions × 11 stimuli × 4 continua + fillers) in four identification tests, each test containing 440 stimuli (20 repetitions × 11 stimuli × 1 continuum + fillers). The stimuli in each test were in a random order and listeners should finish tests one by one.

D. Data Analysis

By combining data from each listener's proportion of /ɪ/ responses to four synthesized continua, the mixed-effects logistic regression models [18] were constructed to explore the relationship between two acoustic cues and the proportion of responses in two groups respectively. The dependent variables were the proportion of English /ɪ/ response. The independent variables were two acoustic cues, duration and spectrum. The model included random intercepts and slopes. The estimated regression coefficients were measured to evaluate the slope of the fitted logistic curves, which was an indication of the sharpness of the categorical boundary and the reliance on acoustic cues.

III. RESULTS

A. Accuracies in the word recognition test

Following the procedure introduced in [15], listeners' mean accuracy scores in the English word recognition test were calculated individually and then submitted to a k-means classifier. The 20 listeners were divided into a low English proficiency (LEP) group and a high English proficiency (HEP) group by the k-means classifier. The LEP group comprised 8 listeners with a low identification accuracy rate (M = 57.86%, SD = 0.07), and the HEP group consisted of 12 listeners with a relatively high identification accuracy rate (M = 79.68%, SD = 0.06). The two groups were significantly different in accuracy scores [$t(18)= -7.761, p<0.001$]. Fig. 2 demonstrates the identification accuracies of the LEP group and the HEP group for both /ɪ/ and /i:/ tokens in the word identification test. Statistical analysis showed that for /ɪ/ tokens, no significant difference was found between the LEP group (56.88%) and the HEP group (73.68%) [$t(18)= -1.838$, $p= 0.099>0.05$], however, for /i:/ tokens, the identification accuracy of the LEP group (58.85%) was significantly lower than that of the HEP group (85.69%) [$t(18)= -4.02$, $p<0.05$]. Further statistical analysis shows that, there was no significant difference of recognition accuracy between /ɪ/ and /i:/ for the LEP group [$t(14)= -0.174$, $p=0.864>0.05$], but a significant difference for the HEP group [$t(22)= -3.039$, $p<0.05$]. This result indicated that listeners in the HEP group had a relatively better ability in identifying English /ɪ/ and /i:/ than the LEP group.

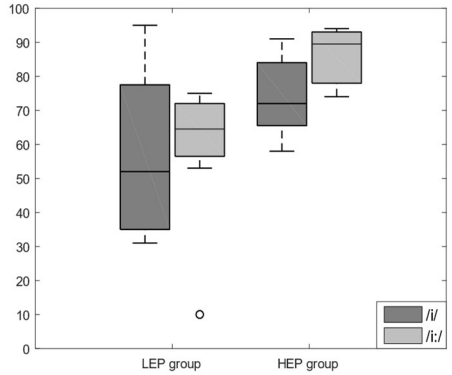
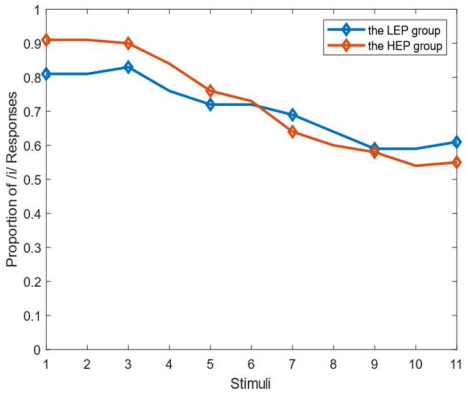


Fig. 2. The mean identification accuracy of words with English /ɪ/-/i:/ between the two groups

B. Perception of durational continua

HEP and LEP groups' perception performances for the two durational continua are shown in Fig. 3. The group-averaged proportion of /ɪ/ responses for each of the 11 stimuli in the /ɪ/ → /i:/ and /i:/ → /ɪ/ continuum are shown in the upper and lower panels of Fig. 3 respectively. For the /ɪ/ → /i:/ continuum (upper panel), neither the HEP group nor the LEP group showed a clear "Z" shape categorical perception curve. However, a general trend that with the increasing vowel duration, listeners' proportion of /ɪ/ responses decreased can still be observed for both groups. It can also be seen in the upper panel of Fig. 3 that the HEP group had a relatively steeper drop of the proportion of /ɪ/ response than the LEP group. The mixed-effects logistic regression models confirmed that vowel duration had a significant effect on the HEP group's vowel identification ($\beta_{HEP}= -0.239$, $p<0.001$) but not on the LEP group's ($\beta_{LEP}= -0.126$, $p=0.101>0.05$), indicating the HEP group might be more sensitive to the durational cue for the English /ɪ-i:/ contrast than the LEP group.

For the /i:/ → /ɪ/ continuum (lower panel of Fig. 3), both the HEP and LEP groups had much lower /ɪ/ responses, which might be due to the fact that the /i:/ → /ɪ/ continuum was synthesized based on vowel /i:/. Again although no categorical perception curve can be observed, both groups' /ɪ/ response decreased as when the vowel duration increased. Statistical analysis of the mixed-effects logistic regression models showed that vowel duration had a significant effect on the HEP group's vowel identification ($\beta_{HEP}= -0.122$, $p<0.05$) but not on the LEP group's ($\beta_{LEP}= -0.1$, $p=0.17>0.05$), indicating the HEP group might rely more on the durational cue than the LEP group.



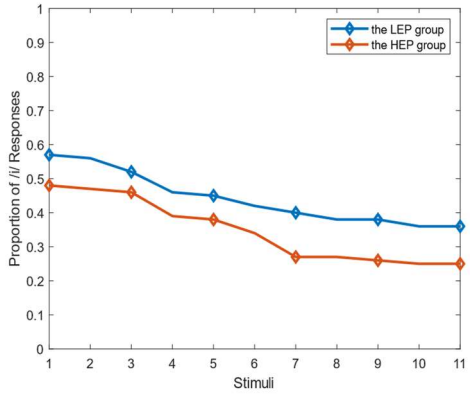
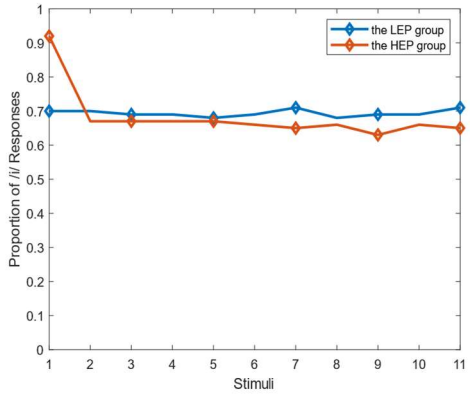


Fig. 3. Perception of durational continua for HEP and LEP groups. The y-axis shows the proportion of English /ɪ/ responses to the 11 stimuli (x-axis) in the /ɪ/ → /i:/ (upper) and /i:/ → /ɪ/ (lower) continua. Note that stimulus 1 always represents the stimulus with the same duration as original /ɪ/ and stimulus 11 is the stimulus with the same duration as original /i:/.

C. *Perception of spectral continua*

Fig. 4 demonstrates the perception performances for the two spectral continua from the HEP and LEP groups. The upper panel of Fig. 4 shows the group-averaged proportion of /ɪ/ responses for each of the 11 stimuli in the /ɪ/ → /i:/ continuum while the lower panel shows the /ɪ/ responses in the /i:/ → /ɪ/ continuum. For the /ɪ/ → /i:/ continuum (upper panel), both HEP and LEP groups had very flat identification curves, indicating both groups possibly were not quite sensitive to the change of spectral cues. The mixed-effects logistic regression models confirmed that spectral cues did not have a significant effect on the HEP group's (β_{HEP}= -0.065, p>0.05) and the LEP group's (β_{LEP}= 0, p>0.05) vowel identification.

For the /i:/ → /ɪ/ continuum (lower panel of Fig. 4), a similar flat identification curve can be seen for the LEP group, and the statistical analysis of the mixed-effects logistic regression model confirmed that spectral cues had no significant effect on LEP groups' vowel identification (β_{LEP}= 0, p>0.05). However, for the HEP group, a large drop of /ɪ/ response can be observed close to the endpoint of the continuum (stimuli 10 and 11). Mixed-effects logistic regression model also showed that spectral cues had a significant effect on the HEP group's vowel identification (β_{HEP}= -0.016, p<0.05). This result suggested the HEP group was more sensitive to spectral cues than the LEP group and demonstrated a more categorical perception like behavior for the /i:/ → /ɪ/ continuum.



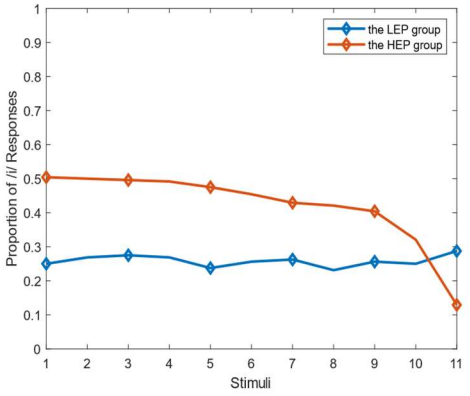


Fig. 4. Perception of spectral continua for HEP and LEP groups. The y-axis presents the proportion of English /ɪ/ responses to the 11 stimuli (x-axis) in the /ɪ/ → /i:/ (upper) and /i:/ → /ɪ/ (lower) continua.

IV. Discussion

The current study investigated the utilizing of durational and spectral cues in the perception of English tense-lax vowel contrast /ɪ-i:/ by native Chinese listeners. A word recognition test with naturally produced stimuli and four vowel identification tests with synthetic vowel continua were conducted. A series of analyses demonstrated that listeners in the HEP group were more sensitive to durational cues and spectral cues than the LEP group. In other words, compared to the LEP group, the HEP group relied more on durational cues and spectral cues.

Previous studies demonstrated that the perception of vowels was categorical [6],[7]. In this study, for the two groups, no categorical perception curve could be seen in the durational dimension, while the HEP group showed a more categorical perception in the spectral dimension than the LEP group. According to theoretical models in the field of the second language (L2) sound acquisition, PAM [19], SLM [20] and NLM [21] all claim that listeners tend to process L2 sound via their first language category. If a L2 sound is close to a L1 category, listeners may ignore the subtle difference between the two language sounds and regard it as a sound in the L1 category. Therefore, in this study, results show that Chinese listeners tend to confuse English /ɪ/ and /i:/ with Chinese /ɪ/ and they still have many difficulties in identifying English /ɪ/ and /i:/, in line with a more continuous representation of vowels [22].

Listeners of English mainly identified /ɪ/ and /i:/ in the duration dimension [11]. In fact, this result was not surprising, as the importance of duration in discriminating /ɪ/ and /i:/ was stressed in the process of L2 acquisition [4]. Therefore, Chinese listeners tend to rely heavily on the durational cues they are more familiar with to identify and discriminate the English tense-lax vowel contrasts. However, an interesting result found in the current study was that the LEP group had no significant reliance on durational cues, which might be due to the random responses of listeners with low proficiency, while for the HEP group, durational cue plays a crucial role in identifying English /ɪ/ and /i:/ [8], [9], [10]. In line with previous studies [13], [14], significant differences between groups were found for native Chinese listeners. In the current study, the result demonstrated that listeners with higher English proficiency also attended to spectral cues more sensitively than listeners with low English proficiency, which means that the cue-weighting strategies changed in the two

groups. Different from [12], English proficiency may shift L2 listeners' cue-weighting strategies from less reliance on spectral and durational cues to heavy reliance. It might be due to different materials used in researches. In fact, duration serves only as a secondary distinguishing feature for English tense-lax contrasts, and sometimes it is not stable and reliable [23]. Therefore, except for durational cue, listeners with high English proficiency also focus more on other acoustic cues, such as spectral cue, to perceive speech sounds.

V. CONCLUSION

In general, the current study investigated the influence of L2 proficiency on the perception of English /ɪ/ and /i:/ by native Chinese listeners. A better acquisition of durational cues and spectral cues to identify /ɪ/ and /i:/ was found in the HEP group than in the LEP group.

ACKNOWLEDGMENT

This study was supported by the grant Scientific and Technological Innovation Team of Jiangsu University of Science and Technology (2020).

REFERENCES

[1] J. Gong, Z. Yang, W. Bellamy, F. Wang, and X. Ji, "Effect of Perceptual Training on the Production of English Tense-Lax Vowels by Native Chinese Speakers," in *2020 International Conference on Asian Language Processing (IALP)*, pp. 115–120, Dec. 2020,

[2] 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.

[3] 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)*, pp. 365–369, Nov. 2017.

[4] Rato, "Effects of Perceptual Training on the Identification of English Vowels by Native Speakers of European Portuguese," in *2014 Proceedings of the International Symposium on the Acquisition of Second Language Speech*, pp. 529–546, 2014.

[5] S. Harnad, "Categorical perception," *Encyclopedia of Cognitive science*, vol. 67, no. 4, 2003.

[6] C. F. Altmann, M. Uesaki, K. Ono, M. Matsuhashi, T. Mima, and H. Fukuyama, "Categorical speech perception during active discrimination of consonants and vowels," *Neuropsychologia*, vol. 64, pp. 13–23, Nov. 2014.

[7] H. Zhang, F. Chen, N. Yan, L. Wang, F. Shi, and M. L. Ng, "The Influence of Language Experience on the Categorical Perception of Vowels: Evidence from Mandarin and Korean," *Proc. Interspeech*, pp. 873–877, Sep. 2016.

[8] J. Gong, D. Xue, W. Bellamy, F. Wang, and X. Ji, "Temporal and Formant Trajectory Analysis of English Tense-Lax Vowels Produced by Native Chinese Speakers," In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 589–594, Dec. 2020.

[9] G. S. Morrison, "L1-Spanish Speakers' Acquisition of the English /i /—/ɪ/ Contrast: Duration-based Perception is Not the Initial Developmental Stage," *Language and Speech*, vol. 51, no. 4, pp. 285–315, Dec. 2008.

[10] Grenon, M. Kubota, and C. Sheppard, "The creation of a new vowel category by adult listeners after adaptive phonetic training," *Journal of Phonetics*, vol. 72, pp. 17–34, Jan. 2019.

[11] D. Kim, M. Clayards, and H. Goad, "A longitudinal study of individual differences in the acquisition of new vowel contrasts," *Journal of Phonetics*, vol. 67, pp. 1–20, Mar. 2018.

[12] W. Hu et al., "Shifting Perceptual Weights in L2 Vowel Identification after Training," *PLOS ONE*, vol. 11, no. 9, Sep. 2016.

[13] E. J. Kong and I. H. Yoon, "L2 Proficiency Effect on the Acoustic Cue-Weighting Pattern by Korean L2 Listeners of English: Production and Perception of English Stops," *Phonetics and Speech Sciences*, vol. 5, no. 4, pp. 81–90, Dec. 2013.

[14] E. J. Kong and S. Kang, "Individual Differences in Categorical Judgment of L2 Stops: A Link to Proficiency and Acoustic Cue-Weighting," *Language and Speech*, vol. 66, no. 2, pp. 354–380, 2023.

[15] C. Prianto, R. Nuraini, and A. T. Wali, "Implementation of K-Means Methods In Clustering Students Ability Levels in English Language," *IJICS Int. J. Inform. Comput. Sci.*, vol. 3, no. 2, pp. 49-58, 2019.

[16] Jie Ji and Aitong Jiang, "The Influence of Dialect on the Perception and Production of Lax-Tense Vowel Distinction in English Learning," *Int. J. Lang. Linguist.*, vol. 10, no. 2, pp. 103–110, Mar. 2022.

[17] Schneider, W., Eschman, A., & Zuccolotto, A. *E-Prime reference guide*. Psychology Software Tools, Incorporated. 2002.

[18] E. J. Kong and J. Edwards, "Individual differences in L2 learners' perceptual cue weighting patterns," in *Proceedings of ICPhS*, 2015.

[19] 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.)," Baltimore: York Press, pp. 171–204, 1995.

[20] J. E. Flege, "Second-language speech learning: Theory, findings, and problems," *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*, Baltimore: York Press, pp. 233–277, 1995.

[21] 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)," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 363, no. 1493, pp. 979–1000, Mar. 2008.

[22] Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. "The Identification and Discrimination of Synthetic Vowels," *Language and Speech*, vol. 5, no. 4, pp. 171–189, 1962.

[23] Y. Jia, and Y, Wang, "Typology of English monophthongs by EFL listeners from Wu dialectal region-A case study of Ningbo and Shanghai," in *Proc. 9th International Conference on Speech Prosody 2018*, pp. 843-847, 2018.