

THE ROLE OF HIGH VARIABILITY PHONETIC TRAINING ON CHINESE EFL LEARNERS' PERCEPTION OF ENGLISH VOWELS IN NOISY ENVIRONMENT

*Qianxi Yu, Ping Tang**

School of Foreign Studies, Nanjing University of Science and Technology, China

ABSTRACT

High variability phonetic training (HVPT) can improve non-native vowel perception. However, most HVPT studies were conducted in quiet conditions, and it remains unclear whether the training effect was generalizable to noisy environments. This study investigated the effect of HVPT on Chinese EFL learners' identification and discrimination of two challenging English tense-lax vowel contrasts /ɑ/-/ʌ/ and /i/-/ɪ/ in noisy conditions. Twenty-four Chinese college students participated in this study. A pretest-training-posttest paradigm was adopted. Participants were trained with English vowel contrasts produced by 10 native speakers of American English using an identification task. Half participants were tested in quiet conditions and the other half in noisy conditions using identification and discrimination tasks. The results showed that listeners' vowel perception performance in both quiet and noisy conditions improved after training, though the improvement varied across vowels. These results thus have implications for HVPT in daily speech communication environments with noise.

Index Terms— High variability phonetic training, L2 speech perception, English tense-lax vowels, Chinese EFL learners

1. INTRODUCTION

Correctly perceiving speech sounds in a second language (L2) can be challenging for L2 learners. Previous studies have reported that Chinese EFL learners (English as a foreign language) showed difficulties in identifying and discriminating between some English vowel contrasts, especially the tense-lax vowel contrasts /ɑ/-/ʌ/ and /i/-/ɪ/ [1-3].

Phonetic training has been widely adopted to address challenges in L2 sound perception. Among various training paradigms, the high variability phonetic training (HVPT), which uses speech stimuli in various speech contexts naturally produced by multiple native speakers with timely feedback, has received particular attention. HVPT has been adopted for its strength in promoting robust L2 learning and generalization to novel speakers and tokens [4-5].

The efficacy of HVPT in improving non-native sound perception has been demonstrated in listeners with various L1 backgrounds, including Chinese EFL learners. However, most HVPT studies focused on the effect of long-term training with the training period varying from weeks to months [6-7]. Only one study [8] has examined the effectiveness of short-term HVPT within 80 minutes and reported minor improvements in identification of /i/-/ɪ/ and generalization of training effects to novel speakers and novel tokens. Therefore, the potential and strength of short-term HVPT need more empirical support.

In addition, most studies examined the training effect in quiet laboratory conditions. However, noise cannot be avoided in daily speech communication. Studies have revealed that non-native listeners can perform as comparatively well as native listeners under quiet conditions, while they tend to perform worse than native listeners when tested in their L2 in adverse listening conditions [9-10]. This raises questions about whether the effect of phonetic training still holds in noisy conditions.

Therefore, the current study aimed to explore the effect of short-term HVPT on Chinese EFL learners. Specifically, we asked whether HVPT improves Chinese EFL learners' identification and discrimination of two challenging English tense-lax vowel contrasts /ɑ/-/ʌ/ and /i/-/ɪ/ in quiet conditions after the short-term training, as demonstrated by improved performance on stimuli produced by both familiar and novel speakers. If yes, we further asked whether such training effect testified in quiet conditions can generalize to speech perception under noisy listening conditions.

2. METHODS

2.1. Participant

Twenty-four native speakers of Mandarin Chinese (11 females; 13 males) were recruited in the current study. They were all non-English majors from Nanjing University of Science and Technology, China, ranging in age from 18 to 23 years (Mean = 21 years, SD = 3). All participants reported that they had studied English as a foreign language for at least six years. No participants reported difficulties related to hearing and speech.

2.2. Stimuli

Two English tense-lax vowel contrasts /a/-/ʌ/ and /i/-/ɪ/ were selected as the training target. Each contrast consisted of 15 word pairs, contributing to a stimuli corpus of 60 CVC monosyllabic words with a variety of onsets and codas (e.g., bot-but for /a/-/ʌ/; bead-bid for /i/-/ɪ/).

Ten native English speakers (5 females, 5 males) with the General American accent were recruited through the website FREELANCER (<https://www.freelancer.com/>) to produce stimuli in this study. All recording was conducted in a sound-treated booth and the sound was collected at a sampling rate of 44100 kHz with a 16-bit depth.

2.3. Procedure

The current study followed a pretest-training-posttest paradigm. Half participants were tested in quiet conditions (the Quiet group) and the other half were tested in noisy condition with stimuli being superimposed with a 6-speaker babble noise (the Noise group; SNR: 10dB). Both groups were trained in quiet conditions. Within each group, half listeners (6 listeners) were trained on the vowel contrast /a/-/ʌ/, and the other half (6 listeners) on /i/-/ɪ/. Both tests and training were conducted using the PsychoPy software [11], and correct responses in the identification and the discrimination tests were automatically collected by the software.

2.3.1. Pretest/Posttest

Pretest and Posttest were identical tests with Pretest conducted before training and Posttest conducted immediately after training.

Prior to Pretest, participants would hear recordings of two common minimal pairs of the target vowel contrast three times. This was done to familiarize the participants with the pronunciation of target vowels. (e.g., did-/ɪ/; deed-/i/).

For each listener, two native speakers (1 female, 1 male) were randomly selected to produce testing stimuli. One speaker produced both testing and training stimuli (the Familiar speaker) while the other speaker only produced testing stimuli (the Novel speaker). Testing stimuli produced by the Familiar speaker were used to test training effects, and stimuli produced by the Novel speaker were used to test generalization effects to novel speakers.

In Pretest/Posttest, participants first completed a two-alternative forced-choice identification task by identifying the vowel in the heard word between two vowel members of the target vowel contrast (e.g., pressing “P” in the keyboard for the vowel /i/ when the word “bead” being played; pressing “Q” for the vowel /ɪ/ when “bid” being played). Testing materials in the identification test included two sets of words. One word in a minimal pair was randomly selected, contributing to a total of 15 words as one set, and

the counterparts of the selected 15 words as the other set. Each participant was tested on either set of the target vowel contrast. This was done to ensure that participants would hear only one word of each minimal pair. All participants would hear 30 stimuli (15 words × 2 speakers) in their identification tests.

Then participants finished the discrimination task in which they had to choose “same” or “different” for two heard sounds (e.g., pressing “P” for “same” when “bead” and “bead” being played; pressing “Q” for “different” when “bead” and “bid” being played). Both two words in every minimal pair of the vowel contrast /a/-/ʌ/ or /i/-/ɪ/ were produced by the two speakers, contributing to 60 stimuli (30 words × 2 speakers) in the discrimination test for each listener.

2.3.2. Training

In the training phase, only the identification task was exploited. The procedure was the same as in the identification test, but feedback in texts of “Correct!” or “Oops, that was wrong!” was provided immediately after each training trial. For a correct answer, training would proceed automatically to the next trial. However, if the answer were wrong, the stimuli in this trial would be played again before the next trial, aiming to familiarize listeners with the pronunciation of the target vowel.

Training stimuli presented to each participant were recordings of all the words in 15 minimal pairs of /a/-/ʌ/ or /i/-/ɪ/. Five speakers out of the speaker corpus with the additional Familiar speaker in Pretest produced training stimuli of high variability. Therefore, a total of 180 stimuli (30 tokens × 6 speakers) served as training materials for each participant. All training stimuli were assigned to three training blocks with 60 trials (30 tokens × 2 speakers) in each block. With a one-minute rest interval between every two training blocks, all participants completed the training within 40 minutes.

2.4. Data analysis

All data analyses were performed in R. The mixed-effects logistic regression model was conducted using the “glmer” function in the R package “lme4” [12], and the “anova” function was used to compare models to calculate the p-value and the Chi-square value.

Tukey-HSD post-hoc tests were performed when a significant main effect or a significant interaction effect was observed, using the “lsmeans” function in the R package “emmeans” [13].

3. RESULTS

3.1. Identification tests

Figure 1 illustrates the accuracy of stimuli produced by familiar and novel speakers in the identification task in

Pretest (before training) and Posttest (immediately after training) for the Quiet group, and accuracy for the Noise group was demonstrated in Figure 2.

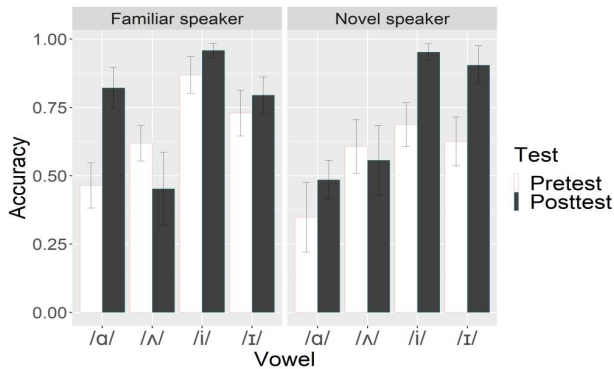


Figure 1. Mean identification accuracy of stimuli produced by familiar and novel speakers in Pretest and Posttest for the Quiet group.

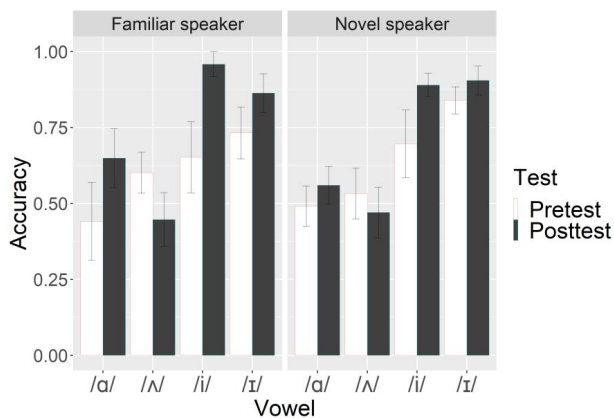


Figure 2. Mean identification accuracy of stimuli produced by familiar and novel speakers in Pretest and Posttest for the Noise group.

A mixed-effects logistic regression model was conducted on the accuracy data of identification tests respectively for the Quiet group and the Noise group (1 = correct response, 0 = incorrect response), with three fixed factors “Test” (2 levels: Posttest vs. Pretest), “Vowel” (4 levels: /a/, /ɔ/, /i/, /ɪ/), “Speaker” (2 levels: Familiar speaker vs. Novel speaker), and two random factors “Participants” (24 participants) and “Item” (15 items).

For the Quiet group, the regression model reported a significant three-way interaction of “Test × Vowel × Speaker” ($\chi^2 = 16.934$, $p = 0.009$), indicating that the significant difference between identification accuracy in Pretest and that in Posttest differs across vowels and speakers.

A Tukey-HSD post-hoc test was performed for this three-way interaction to compare the accuracy of stimuli

produced by the familiar speaker and the novel speaker in Pretest and Posttest for each vowel. The results showed that the accuracy of /a/ revealed a significant Pretest-to-Posttest increase after training for stimuli produced by familiar speakers (contrast: Pretest-Posttest, $\beta = -1.956$, $SE = 0.545$, $z = -3.59$, $p < 0.001$), suggesting that short-term HVPT was effective in improving the identification of /a/ within just one session. In addition, the results also showed that only /i/ (contrast: Pretest-Posttest, $\beta = -2.477$, $SE = 0.817$, $z = -3.031$, $p = 0.002$) and /ɪ/ (contrast: Pretest-Posttest, $\beta = -2.049$, $SE = 0.644$, $z = -3.181$, $p = 0.002$) revealed a significant increase from Pretest to Posttest in the accuracy of stimuli produced by novel speakers, indicating for /i/ and /ɪ/, the short-term HVPT was effective in improving the identification of stimuli produced by novel speakers. This confirmed the effectiveness of HVPT in improving the identification of /i/ and /ɪ/ and its generalizability to novel speakers in quiet conditions.

For the Noise group, the regression model reported a significant two-way interaction of “Test × Vowel” ($\chi^2 = 21.466$, $p < 0.001$), indicating that the differences between the identification accuracy in Pretest and that in Posttest for the Noise group varied across vowels. A Tukey-HSD post-hoc test was then performed for this two-way interaction to compare the accuracy of identification tests in Pretest and Posttest for each vowel. The post-hoc analysis found that only the vowel /i/ revealed a significant increase in the identification accuracy from Pretest to Posttest in noisy conditions (contrast: Pretest-Posttest, $\beta = -2.040$, $SE = 0.503$, $z = -4.055$, $p < 0.001$).

To sum up, HVPT was effective in improving participants’ identification of /a/, /i/, and /ɪ/ in quiet conditions in just one training session. Meanwhile, the training effect of HVPT was generalized to novel speakers for /i/ and /ɪ/. Furthermore, the effect of HVPT in improving the identification of /i/ in quiet conditions still held in noisy conditions, and this improvement was also generalized to the novel speaker.

3.2. Discrimination tests

Figure 3 illustrates the mean accuracy of stimuli produced by three types of speakers in discrimination tests in Pretest (before training) and Posttest (immediately after training) for the Quiet group, and Figure 4 illustrates the same content for the Noise group.

The discrimination test included three types of speakers: (1) Familiar speaker-Familiar speaker: two stimuli in each discrimination trial were produced by the Familiar speaker; (2) Familiar speaker-Novel speaker: one stimulus in each discrimination trial was produced by the Familiar speaker, and the other by the Novel speaker; (3) Novel speaker-Novel speaker: two stimuli in each discrimination trial were produced by the Novel speaker.

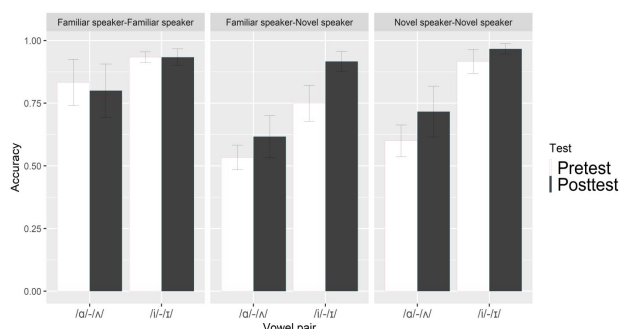


Figure 3. Mean discrimination accuracy of stimuli produced by familiar and novel speakers in Pretest and Posttest for the Quiet group

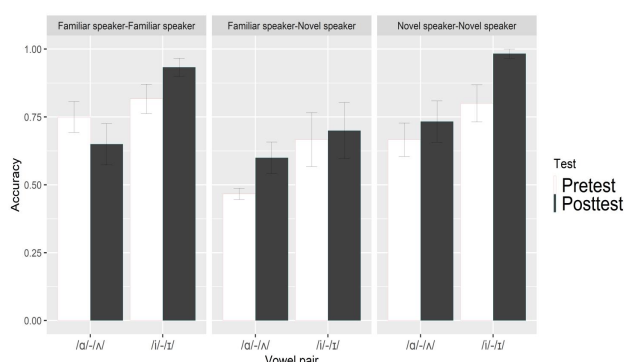


Figure 3. Mean discrimination accuracy of stimuli produced by familiar and novel speakers in Pretest and Posttest for the Noise group

A mixed-effects logistic regression model was respectively performed on the discrimination accuracy data for the Quiet group and the Noise group (1 = correct, 0 = incorrect), with three fixed factors “Test” (2 levels: Pretest vs. Posttest), “Vowel pair” (2 levels: /a/-/ʌ/ vs. /i/-/ɪ/), “Speaker” (3 levels: Familiar speaker-Familiar speaker, Familiar speaker-Novels speaker, Novels speaker-Novels speaker) and three random factors “Participants” (24 participants), “Item1” (15 words for the first stimuli in each discrimination trial), “Item2” (15 words for the second stimuli in each discrimination trial).

For the Quiet group, the regression model reported a significant main effect of “Test” ($\chi^2 = 6.056$, $p = 0.014$), indicating that the accuracy of the discrimination task in Pretest for the Quiet group was significantly different from that in Posttest. A Tukey-HSD post-hoc test was performed for the main effect of “Test”, to compare the accuracy of Pretest and that of Posttest for all vowel contrasts in the current study. The post-hoc analysis found that the accuracy of the discrimination test showed a significant Pretest-to-Posttest increase (contrast: Pretest-Posttest, $\beta = -0.565$, $SE = 0.257$, $z = -2.20$, $p = 0.028$), indicating that HVPT using the identification task was effective in improving participants’ discrimination of /a/-/ʌ/ and /i/-/ɪ/ in quiet conditions, and the improvement also generalized to novel speakers.

For the Noise group, the regression model also found a significant main effect of “Test” ($\chi^2 = 5.569$, $p = 0.018$), indicating the accuracy of discrimination tests in Pretest and Posttest was significantly different. A Tukey-HSD post-hoc test was performed for the main effect of “Test”, and analysis found the discrimination accuracy in Posttest was significantly higher than that in Pretest (contrast: Pretest-Posttest, $\beta = -0.76$, $SE = 0.244$, $z = -3.114$, $p = 0.0018$). Besides, a significant three-way interaction of “Test \times Vowel pair \times Speaker” was observed ($\chi^2 = 11.606$, $p = 0.021$), indicating the accuracy difference in Pretest and Posttest varied across vowel pairs and speaker types. A Tukey-HSD post-hoc test was performed for this three-way interaction. The post-hoc analysis revealed that only the vowel pair /i/-/ɪ/ showed a significant Pretest-to-Posttest increase for the accuracy of stimuli produced by novel speakers (contrast: Pretest-Posttest, $\beta = -2.76$, $SE = 1.046$, $z = -2.649$, $p = 0.0081$). This indicates that HVPT was effective in improving the discrimination of vowel contrasts /a/-/ʌ/ and /i/-/ɪ/, but the improvement only generalized to novel speakers in the case of /i/-/ɪ/. All this confirmed the effectiveness of HVPT using the identification task in improving participants’ discrimination of English tense-lax vowel contrasts /a/-/ʌ/ and /i/-/ɪ/ and its generalization ability for /i/-/ɪ/ to novel speakers in noisy conditions.

To sum up, HVPT was effective in improving the identification of /a/, /i/, /ɪ/, and this improvement on /i/, /ɪ/ generalized to novel speakers in quiet conditions. Besides, HVPT using the identification task also promoted participants’ discrimination of /a/-/ʌ/ and /i/-/ɪ/, and this promotion still held for stimuli produced by novel speakers.

In addition, HVPT was effective in improving the identification of /i/ in noisy conditions, and this improvement was generalized to novel speakers within just one training session. Meanwhile, the discrimination for the Noise group of both /a/-/ʌ/ and /i/-/ɪ/ also improved significantly after training, with the improvement on /i/-/ɪ/ generalized to novel speakers.

4. DISCUSSION

The current study firstly explored whether HVPT can improve Chinese EFL learners’ identification and discrimination of tense-lax vowel contrasts /a/-/ʌ/ and /i/-/ɪ/ in quiet testing conditions after a short-term training. The results showed that HVPT was effective in improving Chinese college students’ identification of /a/, /i/, /ɪ/ and discrimination of both /a/-/ʌ/ and /i/-/ɪ/ in quiet conditions, which is consistent with findings of previous studies that examined the effect of HVPT on improving the perception of English vowels for English learners. Besides, support for the generalizability of HVPT to novel speakers has been found, as reflected in the improved identification of /i/ and /ɪ/ and the discrimination of /a/-/ʌ/ and /i/-/ɪ/ for stimuli produced by novel speakers in quiet conditions. This finding is consistent with previous literature which has confirmed

that the high variability feature of HVPT did promote the transfer of information obtained through training to novel speakers [14-15].

However, the results showed that the identification of /ʌ/ in quiet conditions did not improve after training. A potential reason for this finding is that the identification of /ʌ/ was initially more difficult. Therefore, it may be difficult to improve the identification of /ʌ/ after just 40-minutes training. In fact, the difficulties in the identification of /ʌ/ have been reported in previous studies. For example, a study revealed that the vowel /ʌ/ is highly confusable with /ɑ/ for Mandarin speakers since Mandarin speakers often misidentified /ʌ/ with /ɑ/, while the case of the identification of /ɑ/ was not the same [16]. Furthermore, it has been revealed that the effectiveness of HVPT would disappear for vowels of high confusability [17]. All this may answer for the absent effect of HVPT on the vowel /ʌ/ since training within 40 minutes conducted in the current study may not be enough to improve the identification of the highly confusable vowel /ʌ/.

We then asked whether the effect of HVPT can be generalized into noisy testing conditions. Our results found that the effect of HVPT in quiet conditions still held in noisy conditions, and the effect varied across vowels. In noisy conditions, HVPT remained effective in improving the identification of /i/, with this improvement being generalized to stimuli produced by novel speakers. However, /ɑ/ and /ɪ/ no longer showed support for the effect of HVPT on vowel identification in noisy conditions. In addition, HVPT based on identification tasks was effective in improving the discrimination of both two vowel contrasts /ɑ/-/ʌ/ and /i/-/ɪ/ in noisy conditions just as in quiet testing conditions. Still, only the vowel contrast /i/-/ɪ/ revealed generalization of the training effect. The improved identification of /i/ and discrimination of both two vowel contrasts in noisy conditions further confirmed the effectiveness of HVPT. And a possible reason for such improvement in noisy conditions may lie in the high variability feature of HVPT that stimuli in various word contexts produced by multiple speakers may facilitate L2 learners to exploit more cues when perceiving L2 sounds and promote the formation of more robust categories of L2 sounds, thus to facilitate L2 learners to “rescue” from the masking effect of noises suffered in adverse listening conditions.

However, the identification of /ʌ/ showed no significant improvement when testing under noisy conditions after training, just as in quiet testing conditions. A possible explanation is that the identification of /ʌ/ was initially difficult even in quiet testing conditions, so the interference from noise would make it more difficult to correctly identify the vowel /ʌ/ in noisy conditions, leading to the unsatisfactory effect of HVPT on the vowel /ʌ/.

As for the disappeared identification improvement of /ɑ/ and /ɪ/ (which have been proved in quiet conditions) in noisy conditions, there are two possible reasons. First,

listeners in the Noise group were tested in noisy environments but trained in quiet conditions, so there is a possibility that useful cues learned in quiet training conditions may be masked in noisy testing conditions, thus leading to unsatisfactory performance in the difficult identification task when noise was involved. Secondly, listeners in both Quiet and Noise groups received training of the same quantity, but the training which promoted robust L2 learning in quiet conditions may not be enough to guarantee the robustness of L2 learning in noisy testing conditions. Besides, the generalization of training effect for the discrimination of /ɑ/-/ʌ/ in noisy conditions also disappeared, a possible explanation is that the cues abstracted from training may not be robust enough for correct discrimination of /ɑ/-/ʌ/ for stimuli produced by novel speakers in noisy conditions since the babble in noisy conditions may mask some useful cues of vowels which had been learned through training. In addition, the discrimination of /ɑ/-/ʌ/ may be initially more difficult than that of /i/-/ɪ/, as evidenced by [2] which revealed that the discrimination of /ɑ/-/ʌ/ was the most challenging for Mandarin speakers among six vowel contrasts including /i/-/ɪ/. Therefore, the same quantity of training may not be enough to facilitate participants to abstract the universal knowledge of /ɑ/-/ʌ/ in the same way as that for /i/-/ɪ/ from stimuli produced by multiple speakers when noise was included in tests.

5. CONCLUSION

The present study demonstrated that the high variability phonetic training (HVPT) could effectively improve the identification and discrimination for Chinese EFL learners of two challenging English tense-lax vowel contrasts /ɑ/-/ʌ/ and /i/-/ɪ/ in noisy listening conditions where a six-talker babble was involved. Besides, even the short-term HVPT completed within 40 minutes can promote generalization to stimuli produced by novel speakers. However, the efficacy of HVPT varied across vowels, and the effectiveness of HVPT on a vowel proved in quiet conditions was even lost in noisy conditions. All this has implications that the effectiveness of HVPT in improving non-native vowel perception should be examined on different vowels, and the interference of noise in daily life should be taken into account in training-related studies, thus to further explore the potential of phonetic training in improving non-native sound perception in real-life situations.

6. REFERENCES

- [1] J. E. Flege, O. S. Bohn, and S. Jang, “Effects of experience on non-native speakers’ production and perception of English vowels,” *Journal of Phonetics*, vol. 25, no. 4, pp. 437-470, 1997.
- [2] G. Jia, W. Strange, Y. Wu, J. Collado, and Q. Guan, “Perception and production of English vowels by Mandarin

- speakers: Age-related differences vary with amount of L2 exposure,” *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1118-1130, 2006.
- [3] S. Lei, “Perceptual assimilation of English vowels by Chinese listeners,” *Linguistics in the Netherlands*, vol. 24, no. 1, pp. 150-161, 2007.
- [4] J. W. S. Wong, “The effects of perceptual and/or productive training on the perception and production of English vowels /ɪ/ and /i:/ by Cantonese ESL learners,” *Proceedings of 14th Annual Conference of the International Speech Communication Association*, pp. 2113-2117, 2013.
- [5] Y. H. S. Chang, Y. Yao, and B. H. Huang, “Effects of linguistic experience on the perception of high variability nonnative tones,” *The Journal of the Acoustical Society of America*, pp. 120-126, 2017.
- [6] C. Aliaga-Garcia, “The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English,” *Doctoral Dissertation of University of Barcelona*, 2017.
- [7] Y. Shinohara, and P. Iverson, “High variability identification and discrimination training for Japanese speakers learning English/ɾ/-/l/,” *Journal of Phonetics*, vol. 66, pp. 242-251, 2018.
- [8] B. Cheng, X. Zhang, S. Fan, and Y. Zhang, “The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study,” *Frontiers in Psychology*, vol. 10, pp. 1178, 2019.
- [9] A. Estrella, “Informational masking of multi-speaker babble in English vowel identification for Spanish-English bilinguals,” *Doctoral Dissertation of University of Texas*, 2016.
- [10] L. Zhong, C. Liu, and S. Tao, “Sentence recognition for native and non-native English listeners in quiet and babble: Effects of contextual cues,” *The Journal of the Acoustical Society of America*, pp. 297-302, 2019.
- [11] J. W. Peirce, J.R. Gray, S. Simpson, M. R. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. Lindeløv, “PsychoPy2: experiments in behavior made easy,” *Behavior Research Methods*, Springer, pp. 195-203, 2019.
- [12] D. Bates, M. Mächler, B. Bolker, and S. Walker, “*R package version lme4: linear mixed-effects models using Eigen and S4*,” *Journal of Statistical Software*, vol.1, no. 7, pp. 1-23, 2014.
- [13] R. V. Lenth, “Least-squares means: the R package lsmeans,” *Journal of Statistical Software*, vol.69, no. 1, pp. 1-33, 2014.
- [14] J. S. Logan, “Methodological issues in training listeners to perceive non-native phonemes,” *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 1991.
- [15] X. Wang, and M. J. Munro, “Computer-based training for learning English vowel contrasts,” *System*, vol. 32, no. 4, pp. 539-552, 2004.
- [16] J. Guan, C. Liu, S. Tao, L. Mi, W. Wang, and Q. Dong, “Vowel identification in temporal-modulated noise for native and non-native listeners: Effect of language experience,” *The Journal of the Acoustical Society of America*, vol. 138, no. 3, pp. 1670-1677, 2015.
- [17] T. Wade, and L. L. Holt, “Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task,” *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2618-2633, 2007.