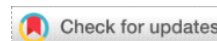




pISSN 2005-8063
eISSN 2586-5854
2026. 9. 30.
Vol.18 No.3
pp. 65-71



Generalization of noise-based high-variability phonetic training in Chinese EFL learners' English vowel perception*

YUYAO QIN**

Department of English Language and Literature, Pusan National University, Pusan, Korea

Abstract

The present study examined the effects of noise-based high-variability phonetic training on Chinese English as a Foreign Language (EFL) learners' perception of English vowels, focusing on whether the training effects were generalized across listening conditions and perceptual tasks. Native Chinese listeners received intensive training to identify English vowels presented as speech-shaped noises. Before and after training, vowel identification in noise was assessed using the vowel reception threshold. Vowel identification in quiet conditions and vowel discrimination were also measured. The results showed significant Group×Test interactions for all three measures, indicating different patterns of change between the training and control groups. The training-related improvement extended from vowel identification in noise to vowel identification in quiet conditions and was also transferred to an untrained discrimination task. These findings suggest that noise-based high-variability phonetic training can enhance L2 vowel perception with benefits generalizing beyond the specific listening conditions and perceptual tasks used during training.

Keywords: phonetic training, speech shaped noise, vowel discrimination, vowel identification, L2 speech perception

1. Introduction

Background noise is a common part of everyday communication and can make speech perception much more difficult. When speech is degraded by noise, listeners need to make use of the acoustic-phonetic cues that remain available while using their linguistic knowledge to understand what they hear (Cooke et al., 2008; Lecumberri & Cooke, 2006). Compared with native listeners, second-language (L2) listeners often find this process more difficult because their L2 phonetic categories are less well established and

they are less able to use the available speech cues effectively (Cooke et al., 2008; Cutler et al., 2008).

Although adult learners generally experience greater difficulty in learning L2 speech sounds, they are still able to improve through perceptual training (Flege, 1995). High-variability phonetic training (HVPT) is characterized by variability in the training input, typically involving multiple talkers and different phonetic contexts. Previous studies have shown that HVPT is effective in improving L2 speech perception, including vowel perception (Lambacher et al., 2005; Lengeris & Hazan, 2010; Mora et al., 2022; Nishi &

* This paper is based on the author's master's thesis.

** 15836084694@163.com, Corresponding author

Received 8 August 2026; Revised 11 September 2026; Accepted 12 September 2026

© Copyright 2026 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kewley-Port, 2007). In the present study, variability was provided not only by multiple talkers and different phonetic contexts, but also by multiple listening conditions. Specifically, the training stimuli were produced by six British English speakers and were presented at four fixed signal-to-noise ratios (SNRs): +3, 0, -3, and -6 dB.

More recent work has examined whether high-variability phonetic training remains effective when target sounds are presented under adverse acoustic conditions. Cooke & García Lecumberri (2018) found that both quiet and noise-based training improved Spanish learners' identification of English consonants, with noise training showing a relative advantage when subsequent testing was also conducted in noise. Mi et al. (2021) similarly reported that training with background noise produced improvements in Chinese listeners' English vowel identification across quiet and noisy test conditions. Training with degraded speech may encourage listeners to make more efficient use of cues that remain informative under masking, rather than relying only on acoustic information that is fully available in quiet (Cooke & García Lecumberri, 2018, 2019). Evidence from adverse-condition high-variability phonetic training also indicates that phonetic learning can generalize beyond the specific training materials (Leong et al., 2018).

Most previous studies have evaluated the effects of noise-based training using identification tasks. Identification performance can be measured by accuracy scores or adaptive threshold measures. In the present study, identification in noise was measured using the vowel reception threshold (VRT), which estimates the signal-to-noise ratio required to achieve a predefined level of vowel-identification performance. Qin & Gong (2023) showed that VRT is sensitive to changes in English vowel perception following noise-based training.

However, both VRT and identification accuracy assess listeners' ability to identify vowel categories, and may not fully capture other aspects of vowel perception. Discrimination therefore provides an additional measure of whether training effects extend beyond vowel identification. Identification and discrimination involve somewhat different perceptual demands. Identification requires listeners to map an acoustic signal onto a phonological category and select the corresponding response label, whereas discrimination only requires them to judge whether two sounds are the same or different and does not necessarily involve explicit category labeling (Iverson et al., 2012). Comparing the two tasks therefore helps determine whether the effects of training are limited to the trained identification task or can also extend to an untrained discrimination task, as training gains may generalize beyond the specific task used during training (Wright & Zhang, 2009). Because the present training involved explicit vowel identification, greater improvement was expected in the identification measures. However, discrimination was also expected to improve if the training strengthened learners' L2 vowel representations beyond the specific task used during training. Accordingly, the present study investigates whether the effects of noise-based vowel identification training extend beyond the conditions and task used during training. Specifically, it examines whether training gains are evident in vowel identification in quiet and whether they also transfer to an untrained vowel discrimination task.

2. Method

2.1. Participant

A cohort of 60 native Chinese participants, consisting of 20 males and 40 females, was recruited from Jiangsu University of Science and Technology. The participants' ages ranged from 19 to 30 years, with a mean age of 23 years. Most participants had been taught English mainly with American English pronunciation models, and none had lived abroad. They were randomly assigned to either the training group (TG; $n=30$) or the control group (CG; $n=30$). All participants reported no problems related to hearing or language, successfully passing a hearing test that required pure-tone thresholds not exceeding 15 dB HL across octave frequencies ranging from 250 to 8,000 Hz. All participants were from regions where the Northern Mandarin dialect is spoken (north of the Yangtze River). All participants were EFL learners with diverse majors, and the majority of them had successfully completed the College English Test Band 4 (CET-4), a standardized English proficiency exam mandated for college students in China, with scores above 425 (out of 710), roughly corresponding to an intermediate level of English ability. Participants took part in the study voluntarily and were rewarded for their involvement.

2.2. Pretest

2.2.1. The English vowel identification test in quiet

In the pretests, participants began by completing the English vowel identification in quiet test, followed by the vowel reception threshold test and the vowel discrimination test. The identification task employed a 15-alternative forced-choice paradigm, requiring participants to assign the auditory stimulus to one of the 15 English vowel categories by clicking the corresponding button on a 3×5 on-screen grid. Fifteen /hVd/ forms (hid, heed, head, had, hudd, hard, hod, hoard, hood, who'd, hade, hide, hoyd, hoad, and how'd), as detailed in Table 1, were displayed on the buttons as labels for the corresponding vowel categories. The set included lexical forms (hid, heed, head, had, hard, hod, hoard, hood, hade, hide, and hoyd), contractions (who'd and how'd), and nonwords (hudd and hoad). The stimuli were produced by three native British English speakers (two males and one female), yielding a total of 270 trials (15 vowels×3 tokens×3 speakers×2 repetitions).

Table 1. Target English vowels in /hVd/ frame words

Vowel category	English vowels	Word
Monophthongs	/i:/	heed
	/ɪ/	hid
	/eɪ/	head
	/æ/	had
	/ɑ:/	hard
	/ʌ/	hudd
	/ɔ:/	hoard
	/ɒ/	hod
	/u:/	who'd
Diphthongs	/ʊ/	hood
	/eɪ/	hade
	/aɪ/	hide
	/ɔɪ/	hoyd
	/əʊ/	hoad
	/aʊ/	how'd

2.2.2. The vowel reception threshold test

During the vowel reception threshold test, hVd stimuli were presented with background noise, and the signal-to-noise ratio for each hVd token was dynamically adjusted based on the participants' previous responses using a 2-down 1-up adaptive procedure, with a fixed step size of 2 dB. For instance, if the current signal-to-noise ratio for a *hid* token was -4 dB and the listener answered incorrectly, the signal-to-noise ratio for the next *hid* token would increase to -2 dB (1-up). If the listener answered correctly, the signal-to-noise ratio for the next *hid* token would remain at -4 dB. If the listener correctly identified the *hid* token at -4 dB again, the signal-to-noise ratio for the third *hid* token would decrease to -6 dB (2-down). Each vowel was presented 18 times, yielding 270 trials in total, and the adaptive track for each vowel terminated after the 18th presentation. Following previous adaptive reception-threshold procedures (Gong et al., 2016), the VRT for each vowel was estimated by averaging the final five SNR values. These final observations represented the later portion of the adaptive track, where the SNR was expected to have converged toward the listener's threshold. Vowel initial SNRs were determined through pilot testing to facilitate convergence of the final SNR values, as shown in Table 2.

During the vowel reception threshold test, the same set of /hVd/ stimuli used in the English vowel identification-in-quiet test was presented with speech-shaped noise. Thus, the only difference between the two pretest tasks was that the vowel reception threshold stimuli were presented under masking conditions. The speech-shaped noise was generated by passing white noise through a 50-coefficient filter shaped according to the long-term average spectrum derived from 200 sentences in a British English corpus (Cooke et al., 2010). All stimuli were normalized to have equal root-mean-square (RMS) energy, and the noise was added immediately before presentation.

Table 2. Vowel-specific initial settings for the vowel reception threshold (VRT) procedure

Vowel Initial VRT (dB)	/ɪ/	/i:/	/e/	/æ/	/ʌ/	/ɑ:/	/ɒ/	/ɔ:/
	-6	-6	-6	-6	-6	-6	-8	-6
Vowel Initial VRT (dB)	/ʊ/	/u:/	/ɛ/	/aɪ/	/ɔɪ/	/əʊ/	/aʊ/	
	-6	-6	-6	-10	-10	-10	-10	

2.2.3. The discrimination test

In the discrimination test, the 15 English vowels (/ɪ, i:, e, æ, ʌ, ɑ:, ɒ, ɔ:, ʊ, u:, ɜ:, eɪ, aɪ, əʊ, aʊ/) were organized into 21 different-vowel contrasts (/ʊ-u:/, /æ-ɑ:/, /æ-ʌ/, /ʌ-ɑ:/, /ɪ-i:/, /ɒ-ɔ:/, /e-eɪ/, /ɔ:-əʊ/, /ɔ:-aʊ/, /ɒ-əʊ/, /əʊ-aʊ/, /ɑ:-ɒ/, /ʌ-ɒ/, /ɜ:-əʊ/, /eɪ-aɪ/, /ɪ-eɪ/, /ɔ:-ɜ:/, /e-æ/, /ʌ-ɜ:/, /ʌ-ʊ/, and /u:-aʊ/) and 15 same-vowel foil categories. The selection of these 21 contrasts was based on the assimilation-overlap patterns reported by Xue (2018), which indicated different degrees of potential perceptual confusability among English vowels for Chinese EFL learners. Ten trials were created for each different-vowel contrast using tokens produced by

the three speakers, resulting in 210 different trials (21 contrasts×10 trials). To ensure an equal number of "same" and "different" trials, 14 same-vowel trials were created for each of the 15 vowel categories, resulting in another 210 trials (15 vowels×14 trials). The discrimination test therefore consisted of 420 trials in total. The subjects were required to determine whether two consecutively played tokens were "same" or "different" (for instance, hud/hoad or had/had) by pressing the corresponding button on the screen during the discrimination test.

2.3. Training

2.3.1. Material

The training materials were different from those used in the pre-test, including several different consonant-vowel-consonant (CVC) frames (e.g., /bVd/, /bVt/, /fVd/). Recordings of these monosyllabic English words were made from another six speakers of British English (4 males, 2 females) who did not appear in the pre-test phase. There were 16 training sessions in total, conducted over four consecutive days with four sessions per day. Each session consisted of 150 trials (15 vowels×10 tokens), resulting in a total of 2,400 training trials (16 sessions×150 trials). The same speech-shaped noise masker used in the vowel reception threshold test was also used during training. However, whereas the SNR was dynamically adjusted in the vowel reception threshold test, four fixed SNR levels were used during training.

2.3.2. Training Procedure

The training group completed 16 sessions over four consecutive days. Each day included four sessions conducted at fixed signal-to-noise ratios of +3, 0, -3, and -6 dB, respectively, as shown in Table 3. Using different signal-to-noise ratio values aimed to promote variability in the availability of speech cues in different noise level, and to simulate everyday noisy environments (Cooke & García Lecumberri, 2018). Similar to the vowel reception threshold test, participants had to classify the vowel they heard in each CVC into one of the 15 categories. Stimulus items were presented in randomized order throughout each training session. Participants completed all 150 trials in each session before proceeding to the next session. Incorrect responses received immediate corrective feedback: the correct response was highlighted, and participants were required to select the correct response and listen to the stimulus again before proceeding to the next trial. Correct responses proceeded directly to the next trial. The TG took part in all stages of the experiment, whereas the CG participated exclusively in the pretest and posttest, without involvement in the training phase.

Table 3. Schedule of the noise-based training (unit: dB)

Training day	Session1	Session2	Session3	Session4
Day 1	3	0	-3	-6
Day 2	3	0	-3	-6
Day 3	3	0	-3	-6
Day 4	3	0	-3	-6

2.4. Posttest

The materials and procedures of the post-test were identical to those of the pretest, with specific details described in Section 2.2

2.5. Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics, Version 26. Three separate 2×2 mixed-design repeated-measures ANOVAs were conducted for the three primary outcome measures: vowel reception threshold, vowel identification accuracy, and vowel discrimination accuracy. For each analysis, test (pretest vs. posttest) was included as the within-subject factor, and Group (TG vs. CG) was included as the between-subject factor. The Group×Test interaction was treated as the primary test of whether pretest-to-posttest changes differed between the TG and CG.

Assumptions relevant to the primary Group×Test interaction were examined using the change-score residuals. Shapiro–Wilk tests showed no significant departures from normality for any of the three outcome measures (all $p \geq .272$), and Levene's tests indicated homogeneity of change-score variances between groups (all $p \geq .629$). No standardized residual exceeded ± 3 . Because Test had only two levels, the assumption of sphericity was automatically satisfied.

Bonferroni-adjusted follow-up comparisons were conducted to examine pretest-to-posttest changes within each group and group differences at each test point. To control for multiple testing across the three primary outcome measures, the p values for the three Group×Test interaction tests were adjusted using the Holm–Bonferroni procedure.

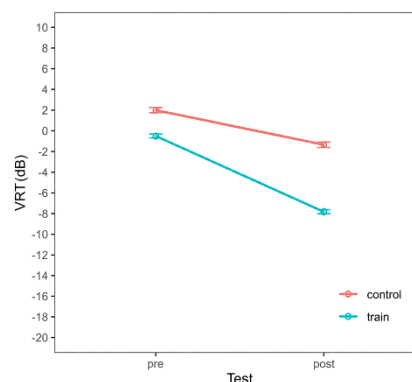
Identification and discrimination accuracy were analyzed as participant-level aggregate scores. This approach was retained because the primary aim was to examine overall pretest-to-posttest changes at the participant level, and all participants completed the same fixed number of trials under the same task conditions. However, aggregation does not retain trial-level variation across individual vowels, contrasts, speakers, or tokens. This limitation is acknowledged in the Discussion.

3. Results

3.1. The Vowel Reception Threshold

3.1.1. Vowel reception thresholds (VRTs) in the pre- and post-test

Figure 1 depicts VRT of English vowels in the Pretest and posttest for the CG and TG. Before training, the CG had a mean VRT of 1.97 dB (95% CI [-0.60, 4.54]), which decreased to -1.36 dB (95% CI [-4.52, 1.79]) in the posttest, indicating a 3.33 dB improvement. The TG had a mean VRT of -0.50 dB (95% CI [-2.54, 1.54]) in the pretest, which decreased to -7.82 dB (95% CI [-9.88, -5.77]) in the posttest, reflecting a 7.32 dB improvement following training. This improvement suggests a significant enhancement in vowel perception accuracy for both groups after the training.

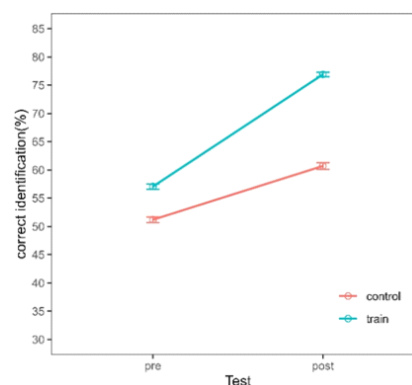


CG, control group; TG, training group.

Figure 1. Vowel reception thresholds of English vowels in the pretest and posttest for the CG and TG.

The Repeated-measures ANOVA revealed significant variations. Specifically, there was a notable difference between the pretest and posttest results, with a statistical significance of $F(1,58)=81.531$, $p<.001$. Furthermore, a significant difference was observed between the CG and the TG, as indicated by the statistical value of $F(1,58)=7.580$, $p<.05$. Additionally, a significant Group×Test interaction was observed, $F(1,58)=11.407$, $p=.001$, $\eta^2=.164$. The interaction remained significant after Holm–Bonferroni correction across the three primary outcome measures (adjusted $p=.003$).

Bonferroni-adjusted follow-up comparisons showed significant pretest-to-posttest reductions in VRT in both the TG and CG (both $p<.001$). The two groups did not differ significantly at pretest ($p=.256$), whereas the TG had a significantly lower VRT than the CG at posttest ($p=.002$).



CG, control group; TG, training group.

Figure 2. pre- and post-test accuracy of CG and TG in identifying English vowels.

3.2. The Identification of English Vowel

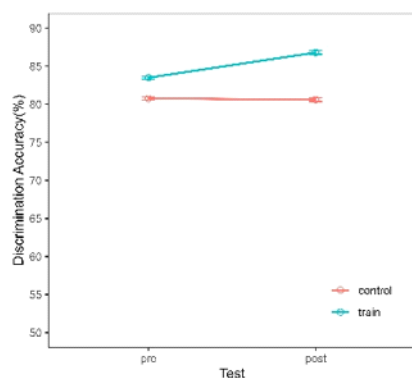
3.2.1. Identification accuracy before and after training

Figure 2 illustrates the changes in identification accuracy for both the CG and TG before and after training. Initially, the CG demonstrated a mean identification accuracy of 51.2% (95% CI [45.78, 56.62]) in the pretest, which subsequently rose to 60.7%

(95% CI [53.82, 67.58]) in the posttest, indicating a notable 9.5% improvement. On the other hand, the TG exhibited a pretest mean identification accuracy of 57.07% (95% CI [51.97, 62.16]), which underwent a substantial enhancement of 19.83% to reach 76.9% (95% CI [72.33, 81.47]) in the posttest.

The results of the Repeated-measures ANOVA indicated that there was a statistically significant difference between the pretest and posttest scores [$F(1,58)=122.887, p<.001$]. Furthermore, a notable difference emerged between the CG and TG [$F(1,58)=9.358, p<.005$]. Additionally, a significant Group $\times$ Test interaction was observed, $F(1, 58)=15.250, p<.001, \eta^2=.208$. The interaction remained significant after Holm–Bonferroni correction across the three primary outcome measures (adjusted $p<.001$).

Bonferroni-adjusted follow-up comparisons showed significant pretest-to-posttest improvements in identification accuracy in both the TG and CG (both $p<.001$). The two groups did not differ significantly at pretest ($p=.224$), whereas the TG showed significantly higher identification accuracy than the CG at posttest ($p<.001$).



CG, control group; TG, training group.

Figure 3. Pre- and post-test discrimination accuracy for the CG and TG.

3.3. The Discrimination Test

3.3.1. Discrimination accuracy of vowel contrasts and Pairs before and after training

Figure 3 depicts the mean discrimination accuracy of all vowel contrasts and pairs for both the CG and TG in the pretest and the posttest. In the pretest, the CG had a mean discrimination accuracy of 80.77% (95% CI [78.42, 83.11]), which decreased to 80.62% (95% CI [77.71, 83.53]) in the posttest, indicating virtually no change. The TG had a mean discrimination accuracy of 83.50% (95% CI [81.41, 85.59]) in the pretest, which increased to 86.8% (95% CI [84.18, 89.42]) in the posttest, reflecting a 3.3% improvement following training. This improvement suggests a slight enhancement in vowel discrimination accuracy for the TG, while the CG did not improve at all.

A repeated-measures ANOVA analysis revealed a statistically significant difference between the pretest and posttest results [$F(1,58)=5.652, p<.05$] and between the CG and the TG [$F(1,58)=7.736, p<.05$]. The interaction effect of test and group was also significant [$F(1,58)=6.754, p<.05, \eta^2=.104$]. The interaction remained significant after Holm–Bonferroni correction across the three primary outcome measures (adjusted $p=.012$).

Bonferroni-adjusted follow-up comparisons showed that discrimination accuracy increased significantly from pretest to posttest in the TG ($p=.004$), whereas no significant change was observed in the CG ($p=1.000$). The two groups did not differ significantly at pretest ($p=.160$), but the TG showed significantly higher discrimination accuracy than the CG at posttest ($p=.004$).

4. Discussion

The present study examined the effects of noise-based high-variability phonetic training on Chinese EFL learners' perception of English vowels using three measures: VRT, vowel identification in quiet, and vowel discrimination. Significant Group $\times$ Test interactions were observed for all three measures, indicating that the pretest-to-posttest changes differed between the TG and CG. These findings suggest that the effects of the HVPT procedure implemented under variable noise conditions extended beyond the specific conditions and task used during training. Specifically, improvement in quiet identification indicates generalization across listening conditions, from noisy training to quiet testing, whereas improvement in discrimination suggests transfer from the trained identification task to an untrained perceptual task. Because the identification and discrimination tasks differed in response format and baseline performance, the numerical gains observed in the two tasks should be interpreted descriptively and should not be taken as evidence that learning was stronger in one task than in the other.

The significant Group $\times$ Test interaction for VRT indicated that the TG and CG showed different pretest-to-posttest changes. The reduction in VRT indicates improved performance on the vowel reception threshold test after training. The control group also showed a decrease in VRT from pretest to posttest, even though they did not receive training. This change may reflect familiarity with the adaptive procedure, repeated exposure to the speech-shaped noise, or other procedural learning effects rather than phonetic training itself (Gong et al., 2016). Importantly, however, the reduction in VRT was significantly larger in the training group than in the control group, supporting a training-related effect relative to the no-training control. This result is consistent with previous studies showing that phonetic training under adverse listening conditions can improve non-native speech perception in noise (Cooke & García Lecumberri, 2018, 2019; Mi et al., 2021). The training exposed participants to speech across multiple SNR conditions. However, because the SNR levels were always presented in the same order, the present design does not allow the effects of particular SNR levels to be separated from possible session-order effects. In addition, because the study did not include a comparable quiet-training group, the specific contribution of noise exposure cannot be distinguished from the overall effect of the HVPT procedure.

The significant Group $\times$ Test interaction for vowel identification indicated greater pretest-to-posttest improvement in the TG than in the CG. This pattern suggests that the effects of the training extended to vowel identification in quiet, a listening condition different from the noisy training condition. This finding is consistent with previous reports showing that training in noise may improve L2 speech perception not only in noise but also in quiet (Cooke & García Lecumberri, 2018, 2019; Mi et al., 2021). However, because the present study did not include a comparable HVPT group trained

in quiet, it cannot determine whether the observed improvement was specifically attributable to noise exposure or whether training under noise provides an advantage over otherwise comparable training in quiet. One possible interpretation is that the observed improvement may reflect stronger mapping between acoustic input and English vowel categories, rather than adaptation to a specific listening condition alone. This interpretation is consistent with previous high-variability phonetic training studies using variability in talkers and phonetic materials (Lengeris & Hazan, 2010; Lengeris & Nicolaidis, 2014; Shinohara & Iverson, 2018). Although the improvement observed in the present study (19.83 percentage points) was smaller than the approximately 48 percentage-point gain reported by Mi et al. (2021), participants in the present study began with considerably higher pretraining accuracy (57.07% vs. approximately 30%), leaving less room for measurable improvement. Differences in the number of vowel categories, training materials, and overall training design may also have contributed to the different magnitudes of learning.

The significant Group×Test interaction for discrimination indicated that the TG and CG changed differently from pretest to posttest. Because the training required participants to identify English vowel categories, whereas the discrimination test required same/different judgments, the significant Group×Test interaction in discrimination is consistent with transfer across perceptual tasks, from the trained identification task to an untrained discrimination task. This finding is consistent with previous studies showing that perceptual learning can transfer across identification and discrimination tasks (Cebrian et al., 2024; Shinohara & Iverson, 2018). At the same time, discrimination accuracy was already relatively high before training (83.50%), leaving less room for further improvement. The same/different format may also have reduced task difficulty and contributed to the high baseline performance.

Overall, the present study showed that the HVPT procedure implemented under variable noise conditions produced training-related improvements in English vowel perception relative to a no-training control, as reflected in VRT, vowel identification in quiet, and vowel discrimination. The results indicate two forms of generalization: first, learning transferred from training in noise to identification in quiet; second, learning extended from the trained identification task to an untrained discrimination task. These findings demonstrate the effectiveness of the present HVPT procedure as implemented under variable noise conditions, but they do not establish a unique advantage of noise-based training over otherwise comparable training in quiet. Several limitations should be acknowledged. First, identification and discrimination were analyzed using participant-level aggregate accuracy scores, which do not directly model trial- or item-level variability. Future studies could use trial-level generalized linear mixed-effects models to account for both participant- and item-level variation. Second, the sample showed a gender imbalance (20 males and 40 females), which may limit the generalizability of the findings. Third, the four SNR conditions were always presented in the same fixed order across the daily training sessions (+3, 0, -3, and -6 dB). Because SNR level was therefore confounded with session position, the effects of particular SNR levels cannot be separated from possible effects of fatigue, reduced attention, adaptation, or procedural learning across sessions. The present design therefore demonstrates exposure to multiple SNR conditions but does not permit an

independent evaluation of the effect of any particular SNR level. Future studies should counterbalance the order of SNR conditions across participants. Fourth, the present study included only an immediate posttest and did not include a delayed posttest. Therefore, the findings demonstrate improvement immediately following training but do not establish whether the observed gains were retained over time. Future studies should include delayed posttests to examine the longer-term retention of training effects.

5. Conclusion

The HVPT procedure implemented under variable noise conditions improved Chinese EFL learners' English vowel perception relative to a no-training control. Significant Group×Test interactions were observed for VRT, vowel identification in quiet, and vowel discrimination, indicating training-related changes across multiple perceptual measures. The improvement in quiet identification suggests that learning generalized beyond the noisy training condition, while the improvement in discrimination indicates transfer from the trained identification task to an untrained perceptual task. These findings highlight the value of using multiple perceptual measures to characterize L2 phonetic learning.

References

- Cebrian, J., Gavalda, N., Gorba, C., & Carlet, A. (2024). Differential effects of identification and discrimination training tasks on L2 vowel identification and discrimination. *Studies in Second Language Acquisition*, 46(4), 1069-1093.
- Cooke, M., & Garcia Lecumberri, M. L. (2018). Effects of exposure to noise during perceptual training of non-native language sounds. *The Journal of the Acoustical Society of America*, 143(5), 2602-2610.
- Cooke, M., & García Lecumberri, M. L. (2019). Non-native consonant acquisition in noise: Effects of exposure/test similarity. *The Journal of the Acoustical Society of America*, 146(1), 297-306.
- Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, 123(1), 414-427.
- Cooke, M., Lecumberri, M. L. G., Scharenborg, O., & van Dommelen, W. A. (2010). Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages. *Speech Communication*, 52(11-12), 954-967.
- Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, 124(2), 1264-1268.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 92(1), 233-277.
- Gong, J., Lecumberri, M. L. G., & Cooke, M. (2016, September). Can intensive exposure to foreign language sounds affect the perception of native sounds? *Proceedings of INTERSPEECH*

- 2016 (pp. 883-887). San Francisco, CA.
- Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160.
- Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2), 227-247.
- Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, 119(4), 2445-2454.
- Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, 128(6), 3757-3768.
- Lengeris, A., & Nicolaidis, K. (2014). Effect of phonetic training on the perception of English consonants by Greek speakers in quiet and noise conditions. *The Journal of the Acoustical Society of America*, 136(4 Suppl), 2107.
- Leong, C. X. R., Price, J. M., Pitchford, N. J., & van Heuven, W. J. B. (2018). High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLOS ONE*, 13(10), e0204888.
- Mi, L., Tao, S., Wang, W., Dong, Q., Dong, B., Li, M., & Liu, C. (2021). Training non-native vowel perception: In quiet or noise. *The Journal of the Acoustical Society of America*, 149(6), 4607-4619.
- Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of non-lexical materials and masking noise. *Phonetica*, 79(1), 1-43.
- Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, 50(6), 1496-1509.
- Qin, Y., & Gong, Y. (2023). Effects of phonetic training on the perception of English vowels in noise by Chinese EFL learners. *Journal of Phonetics*, 98, 101241.
- Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/ - /l/. *Journal of Phonetics*, 66, 242-251.
- Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society B*, 364(1515), 301 - 311.
- Xue, D. (2018). *An experimental study on Chinese EFL learners' perception and production of English vowels* (Master's thesis). Jiangsu University of Science and Technology, Zhenjiang, China.
- **Yuyao Qin**, Corresponding author
Ph.D. Students, Dept. of English Language and Literature
Pusan National University
2, Busandaehak-ro 63 beon-gil, Geumjeong-gu,
Busan, 46241, Korea
Tel: +82-10-2436-2198
Email: 15836084694@163.com
Areas of interest: Phonetics, Phonology, Sociolinguistics