THE INTEGRATION OF ACOUSTIC-CUE VISUALIZATION IN HVPT OF ENGLISH VOWEL PERCEPTION BY CHINESE ESL LEARNERS

Chen Hsueh Chu[a], Han Qianwen[b] and Tian Jingxuan[a]

The Education University of Hong Kong[a], City University of Hong Kong[b]
hsuehchu@eduhk.hk

ABSTRACT

Chinese ESL learners in Hong Kong frequently have pronunciation problems for the vowel pairs /ɪ/-/iː/, /ʊ/-/uː/, /ʌ/-/ɑː/, /e/-/eɪ/, and /aʊ/-/ɑː/. Whether it is also challenging for them to perceive the vowel pairs remains unclear. This study aims to examine how Chinese learners of English in Hong Kong perceive these vowel pairs and investigate the effectiveness of integrating acoustic-cue visualization in the phonetic training of English vowel perception. The study consisted of three interrelated parts: a pre-test, four training sessions, and a post-test. 54 ESL learners whose L1 was Chinese participated in this study. The results show that (i) identifying /ʊ/, /uː/ and /ʌ/, discriminating /ɑː/ from /aʊ/ and /uː/ from /ʊ/ were challenging; (ii) the training improved the subjects' perception of most vowels; and (iii) the improvement was more significant for the subjects with lower English proficiency.

Keywords: English vowels, acoustic phonetics, pedagogy, second language acquisition

1. INTRODUCTION

Perceiving and producing second language (L2) sounds are challenging because of the mismatch in phonetic properties and phonological categories between a first language (L1) and an L2 [1, 2], especially in Hong Kong, where the linguistic situation is complicated. Hong Kong residents usually have Cantonese as their L1, English as an L2, and Mandarin as an L3. Based on [3], adult English learners in Hong Kong have divergent English vowel pronunciation problems, such as confusions in the vowel pairs /ɪ/ and /iː/, /ʊ/ and /uː/, /ʌ/ and /ɑː/, /e/ and /eɪ/, and /aʊ/ and /ɑː/. [4] examined the perception of English consonants and vowels by 40 Cantonese ESL learners. The results indicate that the subjects' perception difficulties do not align with their production problems. The finding highlights the need for further investigation of the perception difficulties and their alignment with the production problems. Furthermore, the possible perception difficulties call for the inclusion of training specifically for improving English vowel perception.

To compensate for the traditional teaching of pronunciation and perception, such as imitation and repetition, researchers have been exploring more effective ways to facilitate phonological acquisition by conducting phonetic training. High variability phonetic training (HVPT), which uses multiple minimal pairs produced by multiple talkers as training materials, could improve the perception of non-native sounds [5, 6]. [7] encourages learners to develop accommodation skills and be exposed to English speakers with different L1s since it is more likely for a learner to communicate with a non-native speaker. In most HVPT practices, the subjects get immediate feedback on the correctness of their judgment after giving a response to each stimulus.

[8] suggests that the effectiveness of phonetic training is not affected by the amount of L2 sound exposure for learners but rather by the extent of learners' attention and L2 sound-processing ability on phonetic cues. The physical representations of the phonetic cues using acoustic tools could enable learners to visualize pronunciation and make modifications. The acoustic cues of pitch height, pitch contour, and duration were used in [9], successfully improving Cantonese speakers' pronunciation of Mandarin tones by visualizing the acoustic cues in Praat. Furthermore, [10] conducted an acoustic spectrographic instruction on the production of English vowel contrasts /i/ and /ɪ/, which improved the subjects' production and perception of the vowels.

This study combines the training methods used in HVPT [6] and the feedback of acoustic-cue visualization in [9, 10]. The subjects were trained by performing auditory judgments in the identification and discrimination tasks after listening to words in minimal pairs uttered by various talkers who were native speakers or learners of English. Immediate feedback was given to the subjects on the correctness of their judgment, and explained why they were correct or incorrect with acoustic-cue visualization after they made each response. The two acoustic cues used in this study were 'vowel length' and 'formant change'. Since the subjects in this study did not have prior knowledge about acoustic phonetics, the acoustic cues used in training were assumed to be relatively simple for the identification or discrimination of two similar vowels, though there may be other differences between the two vowels. Therefore, this paper aims to answer two questions:

a. Which vowels are difficult for the perception by Chinese English learners?
b. Does the use of HVPT with acoustic-cue visualization feedback improve the identification and discrimination of the vowel pairs?

2. METHOD

54 English learners whose L1 was Chinese were involved in this study. They were university students aged from 18 to 25. Based on self-reports of public English exams such as IELTS, GRE, and DSE, 28 subjects had relatively lower English proficiency, and the other 26 had higher English proficiency. All subjects participated in three stages of the experiment – a pre-test, four training sessions, and a post-test.

The pre- and post-tests included a word identification task and a discrimination task, and the rationales of task design were consistent in the pre- and post-tests. Three sets of minimal pair words for each vowel pair /ɪ/-/iː/, /ʊ/-/uː/, /ʌ/-/ɑː/, /e/-/eɪ/, and /aʊ/-/ɑː/ (30 test words in total) were selected to be used as the stimuli. The stimuli were monosyllabic words commonly used in daily communication contexts. The words were read by multiple native speakers from the online dictionaries and learners from [3].

/ɪ/-/iː/		/uː/-/ʊ/	
fist /fɪst/	feast /fiːst/	full /fʊl/	fool /fuːl/
ship /ʃɪp/	sheep /ʃiːp/	foot /fʊt/	fooot* /fuːt/
did /dɪd/	deed /diːd/	look /lʊk/	Luke /luːk/

*pseudo-words produced by the learners [3] were also included as stimuli.

Table 1: Examples of the stimuli

In the identification task, the subjects answered 30 questions (i.e., Which is more likely to be the word read in the audio?) after listening to an audio recording of one stimulus, and identified the word from two minimal pair words. Each question in the identification task contains one test word. In the discrimination task, the subjects answered 45 questions (i.e., Are the two words you just heard [the test word]?) by responding "same" (i.e., Yes, the two words are [the test word]) or "different" (i.e., No, one of the words is [the distractor]) after listening to two audio recordings. Three questions contained each minimal pair. Take the pair "did-deed" for example. The first question contains two recordings of "did", the second question contains two recordings of "deed", and the third question contains a recording of "did" and another recording of "deed". The recordings were of different talkers.

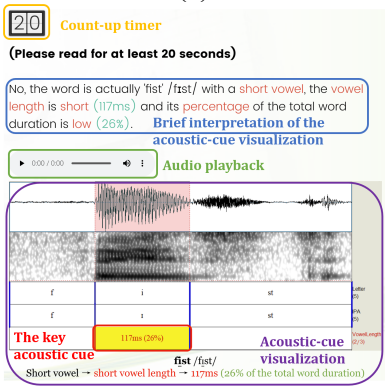
After the pre-test, the subjects received four self-access online training sessions (45 minutes for each session). The training was conducted on Qualtrics. Table 2 presents the topics for the training sessions. The structure for each training session consisted of (i) a 10 to 15-minute instruction video about the acoustic cue and how to connect the acoustic cue with the target vowels; (ii) an identification task with acoustic-cue visualization feedback; and (iii) a discrimination task with acoustic-cue visualization feedback.

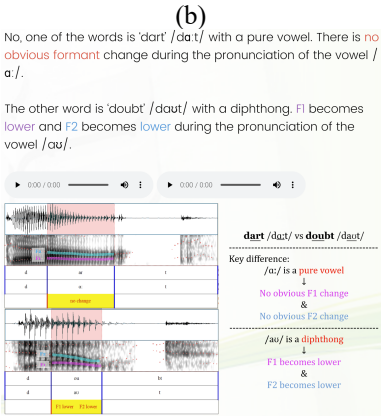
Session	Topics
1	English IPA symbols The acoustic cues and their connection with English sounds
2	The short vowels /ɪ ʊ ʌ/ and the long vowels /iː uː ɑː/ The acoustic cue of '**vowel length**'
3	The pure vowels /e ɑː/ and the diphthongs /eɪ aʊ/ The acoustic cue of '**formant change**'
4	Recap and self-reflection

Table 2: Topics of each training session

The arrangement of the identification and discrimination tasks was consistent with the ones in the pre- and post-tests, except the subjects received immediate acoustic-cue visualization feedback in the training but not in the tests. To complete each identification/discrimination task in the training, the subjects listened to the audio recording(s) and selected the answer within 20 seconds. After giving each response, they would be redirected to a new page with three types of information: (i) the correctness of their perceptual judgment (ii) acoustic-cue visualization feedback and (iii) explanations regarding the connection between the acoustic cue and the target vowels. Since the subjects do not have prior knowledge of acoustic phonetics, the acoustic cues used in this training are simple. Each vowel pair could be differentiated by one type of acoustic cue, namely 'vowel length' or 'formant change'. The acoustic cues are assumed to be the most direct, though there are other differences between the two target vowels.

(a)





Figures 1a-b: Examples of the acoustic-cue visualization feedback for the (a) identification task and (b) the discrimination task

3. RESULTS

Overall, the subjects' identification accuracy rate for most of the target vowels and the discrimination sensitivity for all the target vowels had been improved after receiving the training sessions. The identification accuracy rate refers to the average percentage of correct identification of the stimuli in the identification task across the 54 subjects. The discrimination sensitivity refers to the d prime (*d'*) value based on the Signal Detection Theory [11], by subtracting the z-score of false alarm rate (proportion of "same" questions items to which the subjects responded "different") from the z-score of hit rate (proportion of "different" question items to which the subjects responded "different").

Figure 2 illustrates the identification accuracy rate and the discrimination sensitivity (*d'*) of the subjects in the pre- and post-tests. As /ɑː/ was contrasted with /ʌ/ and /aʊ/ contained by different sets of test words, the one paired with /ʌ/ is presented as /ɑː/, and the other one paired with /aʊ/ is presented as /ɑː/* for differentiation. In the pre-test, the subjects identified /e eɪ aʊ/ with the highest accuracy rates (88% to 91%) and /ʊ uː ʌ/ with the lowest accuracy rates (43% to 53%). In the post-test, similarly, the subjects identified /e eɪ aʊ/ with the highest accuracy rates (90% to 94%). The accuracy rates of identification were the lowest for /uː/ and /iː/ (52% to 58%). Identification accuracy rates of eight vowels were improved. Based on the t-test results, significant improvements were found for /ʊ/ [$t(53) = 4.24$, $p < .001$], /ʌ/ [$t(53) = 3.24$, $p < .001$], and /ɑː/* [$t(53) = 3.01$, $p = .002$]. Furthermore, the training appeared to be more effective in improving identification accuracy for the subjects with lower English proficiency than those with higher proficiency. The identification accuracy of only one vowel /ʊ/ was significantly raised for the subjects with higher proficiency, $t(25) = 2.92$, $p = 0.002$. For the subjects with lower proficiency, significant improvements were found for four vowels /ʊ/ [$t(27) = 3.07$, $p = .001$], /ʌ/ [$t(27) = 2.71$, $p = .004$], /ɑː/ [$t(27) = 2.10$, $p = 0.019$], and /ɑː/* [$t(27) = 3.06$, $p = .001$].

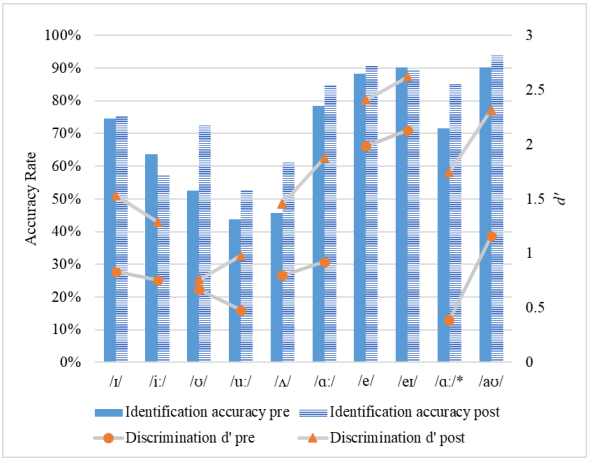


Figure 2: Identification accuracy rate and discrimination sensitivity (*d'*) in the pre- and post-tests

As shown in Figure 2, in the pre-test, the subjects' discrimination sensitivity to /e/ and /eɪ/ was higher than the other vowels, while it was the lowest for /ɑː/* discriminated from /aʊ/ and /uː/. In the post-test, the subjects' discrimination sensitivity was still the highest for /e/ and /eɪ/. In addition, the subjects became more sensitive to /aʊ/ robustly. The sensitivity was the lowest for /ʊ/ and /uː/. The overall discrimination sensitivity to most vowels except for /ʊ/ had been improved. The most remarkable improvement could be observed in the discrimination of the vowel pair /ɑː/*-/aʊ/ from each other, followed by /ɑː/-/ʌ/. Improvements could also be seen for the /ɪ/-/iː/ and /e/-/eɪ/ pairs and /uː/.

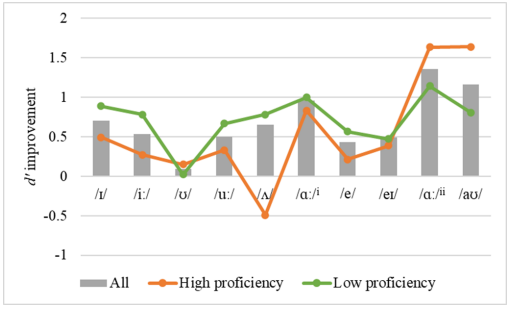


Figure 3: Discrimination sensitivity (*d'*) improvement by the subjects with high or low English proficiency

Comparatively, the improvement on seven of the target vowels in terms of discrimination sensitivity was more salient for the subjects with lower

proficiency than those with higher proficiency. As presented in Figure 3, the *d'* improvement was similar between the two proficiency groups for /ʊ/, /ɑː/, and /eɪ/. The low proficiency group outperformed the high proficiency group in the improvement of five vowels (i.e., /ɪ/, /iː/, /uː/, /ʌ/, and /e/). The high proficiency group outperformed the low proficiency group in improving the discrimination of the /ɑː/*-/aʊ/ pair.

4. DISCUSSION

The vowel pairs difficult for the perception by Chinese ESL learners can be pinpointed from the subjects' performance in the identification and discrimination tasks of the pre-test. Although Chinese ESL learners were found to confuse /e/ and /eɪ/ in their speech production, the identification accuracy (around 90%) and discrimination sensitivity (around 2) were indisputably high in the pre-test. Therefore, the perception of /e/ and /eɪ/ by the subjects in this study was relatively easy. The subjects achieved moderate levels of identification accuracy (above 70%) but relatively low levels of discrimination sensitivity (0.4 to 1.2) for /ɪ/, /ɑː/, /ɑː/* and /aʊ/. It was not difficult for the subjects in this study to identify /ɪ/, /ɑː/ and /aʊ/. However, it was challenging for them to discriminate /ɪ/ from /iː/, /ɑː/ from /ʌ/, and /ɑː/ and /aʊ/ from each other. Both the identification accuracy (43% to 63%) and discrimination sensitivity (from 0.5 to 0.8) were low for /iː/, /ʊ/, /uː/, and /ʌ/. It was the most difficult for the subjects to identify /iː/, /ʊ/, /uː/, and /ʌ/ and discriminate /iː/ from /ɪ/, /ʊ/ and /uː/ from each other, and /ʌ/ from /ɑː/. The patterns of English vowel perception difficulties by Chinese ESL learners in this study are summarized in Figure 4.

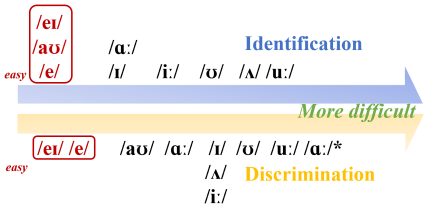


Figure 4: English vowel perception difficulties by Chinese ESL learners

Overall, the use of acoustic-cue visualization in the phonetic training improved the identification and discrimination of the vowel pairs. Among the ten target vowels, improvements in identification accuracy can be observed for eight vowels other than /iː/ and /eɪ/. The identification accuracy improvement of /ʊ/, /ʌ/, and /ɑː/was significant. For discrimination sensitivity, the improvements were apparent for all the target vowels except for /ʊ/. The discrimination sensitivity to the vowel pairs /ɑː/-/aʊ/ and /ɑː/-/ʌ/ was considerably raised. In general, integrating the acoustic cues of 'vowel length' and 'formant change' in the phonetic training effectively improved the perception of English vowels.

The effects of the training varied for subjects with different levels of English proficiency. The low proficiency group significantly raised the identification accuracy of four target vowels, while the high proficiency group significantly raised the accuracy of only one target vowel. Therefore, the training improved the identification accuracy of the subjects with lower English proficiency more effectively. For discrimination sensitivity, the low proficiency group exhibited more enhancement for five vowels, four of which were trained with the integration of the 'vowel length' acoustic cue. In comparison, the high proficiency group showed more improvement for two vowels, which were trained using the 'formant change' acoustic cue. Although the training was effective in improving discrimination sensitivity for both proficiency groups, the acoustic cue of 'vowel length' facilitated the subjects with lower English proficiency more, whereas 'formant change 'was mastered better by the subjects with higher English proficiency. It may be interpreted that the connection between the acoustic cue of 'formant change' and the target vowels was more advanced and required more language knowledge to be processed by the subjects. The connection between the acoustic cue of 'vowel length' was more straightforward and thus easier to be mastered. Chinese ESL teachers should consider learners' proficiency levels when teaching English vowels and select appropriate target features.

5. CONCLUSION

This study explored the perception difficulties in English vowels by Chinese ESL learners in Hong Kong and investigated the impact of integrating acoustic-cue visualization in the phonetic training on the subjects' perception. Identifying /ʊ/, /uː/, and /ʌ/ and discriminating /ɑː/ from /aʊ/ and /uː/ from /ʊ/ were the most difficult for the Chinese ESL subjects in this dataset. Overall, the training improved the identification and discrimination for most of the target vowels. The improvements exhibited by subjects with disparate English proficiency varied. Vowel identification accuracy was improved more effectively for the subjects with lower English proficiency. Although discrimination sensitivity was raised for both proficiency groups, the low proficiency group responded to the acoustic cue of 'vowel length' better, while the high proficiency group internalized the acoustic cue of 'formant change' better. For future studies, comparisons with other teaching approaches can be conducted.

6. REFERENCES

[1] Best, C. T. 1995. A direct realist view of cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Baltimore: York Press, 171-204.

[2] Flege, J. E. 1995. Second language speech learning: theory, finding, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross language Research*. Baltimore: York Press, 233-277.

[3] The spoken English corpus of Chinese and Non-Chinese learners in Hong Kong https://corpus.eduhk.hk/esl_learner_corpus/#/home

[4] Chan, A. Y. 2011. The perception of English speech sounds by Cantonese ESL learners in Hong Kong. *TESOL Quarterly* 45(4), 718-748.

[5] Barriuso, T. A., & Hayes-Harb, R. 2018. High variability phonetic training as a bridge from research to practice. *The CATESOL Journal* 30(1), 177-194.

[6] Thomson, R. I. 2012. Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning* 62(4), 1231-1258.

[7] Jenkins, J. 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics* 23(1), 83-103.

[8] Aliaga-García, C., & Mora, J.C. 2015. Assessing the effects of phonetic training on L2 sound perception and production. *Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*, 11-27.

[9] Chen, H. C., & Han, Q. (2019). The effectiveness of acoustic training on tone acquisition by Hong Kong learners of Mandarin. *Proceedings of the 19th International Congress of Phonetic Sciences* Melbourne, 1957-1961.

[10] Quintana-Lara, M. 2014. Effect of acoustic spectrographic instruction on production of English /i/ and /ɪ/ by Spanish pre-service English teachers. *Computer Assisted Language Learning* 27(3), 207-227.

[11] Abdi, A. (2007) Signal Detection Theory (SDT). In: Salkind, N. (ed), *Encyclopaedia of Measurement and Statistics*. Sage, Thousand Oaks, 886-889.