COMPARING THE PERCEPTUAL TRAINING EFFECTS ON THE PERCEPTION AND PRODUCTION OF ENGLISH HIGH-FRONT AND HIGH-BACK VOWEL CONTRASTS BY CANTONESE ESL LEARNERS

Janice Wing Sze Wong

Hong Kong Baptist University
janicewong@hkbu.edu.hk

ABSTRACT

This study investigated the effects of High Variability Phonetic Training (HVPT) in training both the perception and production of English high-front vowels /ɪ/-/iː/ and high-back vowels /ʊ/-/uː/ by Cantonese ESL learners. Results showed that the HVPT was generally effective in improving the subjects' perception of the two vowel pairs and perceptual learning could generalize to new words and new speakers and be transferred to the production domain, but more perception and production improvement could be observed in the high-front vowel pair than the high-back one. Acoustic analysis also showed that although the subjects tended to exaggerate the vowel duration, the F1 and F2 values also changed, but more was observed in the high-front vowel pair. This is not consistent with SLM as Cantonese has vowels which are similar with the two target L2 vowel pairs but the extent of learning of both target vowel pairs was different.

Keywords: High Variability Phonetic Training, speech perception and production, SLA

1. INTRODUCTION

A number of studies have shown that High Variability Phonetic Training (HVPT) which involves the use of perceptual stimuli under different phonetic environments and produced by various speakers is useful for training L2 learners the perception of L2 phonemes [e.g. 3, 7, 12]. Some also showed that learning in perception can be transferred to the production domain [e.g. 15, 16, 17]. However, most of them focus on training only one pair of non-native contrast [3, 7, 9, 13, 15, 16, 17]; those that report training a greater set of contrasts seldom compare the ease of perception or production of different vowel pairs [8, 11, 12, 14].

The present study aimed to investigate the effect of the HVPT on the perception and production of two English vowel pairs, the high-front /ɪ/-/iː/ and the high-back /ʊ/-/uː/. Although similar vowels (/i/ and /u/ and lowered allophonic [ɪ] and [ʊ] in closed syllables) are present in Cantonese inventory, studies reported that the two English vowel pairs are posing a lot of problems among Hong Kong Cantonese ESL learners [4, 6, 10]. This is well predicted by the Speech Learning Model (SLM, [5]) that L2 phonemes will be more difficult to learn when they are closer to an L2 category. This study also hoped to investigate, though preliminarily, whether there will be any interaction between the two target vowel pairs in the process of learning by comparing the subjects' perception and production performance of the two vowel pairs before and after training.

2. METHOD

2.1 Participants

Thirty-five native Hong Kong Cantonese speakers (average age = 17.13, 19F16M) with English as their L2 were recruited for the study and they were all trained over 20 sessions of HVPT. They all started learning English as an L2 for 14.02 years (SD = .89). They had not resided in any English-speaking countries and reported no history of hearing or speaking impairment.

Eight other native speakers of General American English (4F4M) were also invited to produce all the test and training perceptual stimuli. Their ages ranged from 25 to 45.

2.2 Training

All the subjects took two training sessions per day for a total of 10 days. A 30-minute break was given between sessions. The HVPT was administered in the form of a two-alternative-forced-choice (2AFC) identification task from a computer program designed by the researcher.

2.2.1. HVPT

A total of 80 stimuli produced by six different native English speakers were presented to the subjects with 20 /ɪ/ and 20 /iː/stimuli in Block A and 20 /ʊ/ and 20 /uː/ stimuli in Block B. All the stimuli were one of the counterparts in a minimal word pair contrasting the two vowels (i.e. either “bit” or “bead”). All the stimuli were arranged randomly in each session.

During training, immediate feedback was given; at the end of each session, their total scores were also shown for track-keeping of the training progress. All data were saved into a Microsoft Access database.

2.2.2. Stimulus Materials

All the perceptual training stimuli were made by six of the eight native English speakers. Each speaker produced 40 pairs of /ɪ/-/iː/ and 5 pairs of /ʊ/-/uː/ word pairs but repeated 8 times (due to limited number of high-back vowel minimal pair contrast), contributing a total of 80 stimuli with a wide variety of phonetic environments (with different CVC). The seventh speaker produced the same stimuli for TG3 (familiar words by a new speaker).

Also, one of the above six native speakers, i.e., a familiar speaker to the subjects, also recorded a new word list for TG2 (new words by a familiar speaker) which included 40 /ɪ/-/iː/ and 5 /ʊ/-/uː/ (repeated 8 times) minimal word pairs. The last speaker who had not recorded anything for the training or the tests, known as a new speaker, recorded another new list with both vowel pairs minimal pairs for TG1 (new words by a new speaker).

Each speaker read the tokens at least two times to avoid intra-speaker variability and they appeared in the training program with equal frequencies.

2.3 Pretest/Posttest/Generalization Tests

2.3.1. Production Pre/Posttests

All the subjects took this production test before and after the training. A word list of 30 words (10 /ɪ/, 10 /iː/, 5 /ʊ/, and 5 /uː/) plus 10 distractors was given to them and these tokens might appear in the perception tests or the training. They recorded the words in isolation, one at a time by reading from a screen and into a headset-mounted microphone with Adobe Audition 1.5 software for digitization (sampling rate at 44.1 kHz). They did five practice trials first and were asked to produce the tokens with natural loudness and speaking rate. No audio prompts or instructions were given during the recording. They could also pause and resume the recording based on their own pace. The test could be completed within 15 minutes and was taken before the perception tests to avoid any cueing effects.

2.3.2. Perception Tests

The subjects completed the test within 30 minutes before and after they finished the entire training program. There were 90 questions in total (40 /ɪ/-/iː/ stimuli and 40 /ʊ/-/uː/ stimuli, plus 10 distracters). Five practice trials were completed before the test. Before they confirmed their answers, they listened to the stimuli and chose the answer from three choices with conventional English orthography, or a blank for free answer for them to type their own word. The frequency of occurrence of the correct answer appeared in the four serial positions, i.e. word 1, word 2, word 3 and an open answer, were equal, allowing the chance level to be correctly and fairly inferred at 25%. This design was aimed to reveal a more genuine and reliable performance of the subjects by avoiding the 50% probability of correct answer obtained in typical two-alternative identification test.

The subjects were also given three generalization tests. For Test of Generalization 1 (TG1), the subjects heard 40 /ɪ/-/iː/ stimuli and 40 /ʊ/-/uː/ stimuli. All new words which had not appeared in previous tests or training and they were all produced by a new speaker. For Test of Generalization 2 (TG2), the subjects were given 40 /ɪ/-/iː/ and 40 /ʊ/-/uː/ new stimuli spoken by a familiar speaker. The last generalization test, Test of Generalization 3 (TG3), the subjects were given the same 40 /ɪ/-/iː/ and 40 /ʊ/-/uː/ used in pre/posttest but were all produced by another new voice. The procedures of the three generalization tests were the same as those administered in the perception pretest.

2.4 Data Analysis

All production test tokens were transcribed twice by a phonetically-trained native speaker of General American English studying Applied English Linguistics. The correlation coefficient of the two transcriptions was computed and the intra-rater judgment was obtained at $r = .973$ ($p < .001$).

A follow-up acoustic analysis using the *Praat* speech analysis software on *F1* and *F2 values* (measured at midpoints) and the vowel durations was also conducted [1].

3. RESULTS

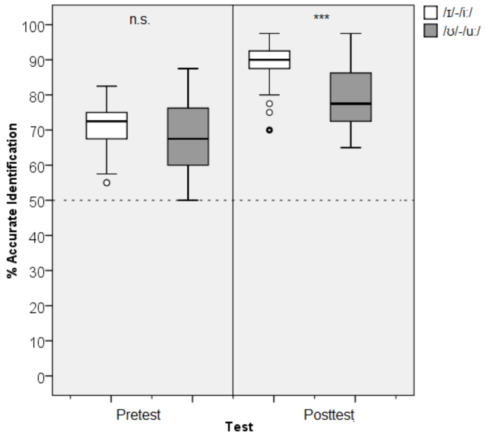
3.1 Perceptual Performance

3.1.1. Pretest vs. Posttests

Figure 1 displays the perceptual learning results of the two vowel pairs by comparing the subjects' pretest and posttest results.

A two-way repeated-measures ANOVA was computed using Test (pretest, posttest) and Vowel (high-front, high-back) as factors. It showed significant main effects of Test [$F(1,34) = 202.81$, p

Figure 1: Mean percentages of accurate identification of the two vowel pairs (high-front = white boxes; high-back = dark boxes) in the pretest (left) and the posttest (right) [*** $p < .001$; n.s. = $p > .05$]. The dashed line indicates the chance level performance.

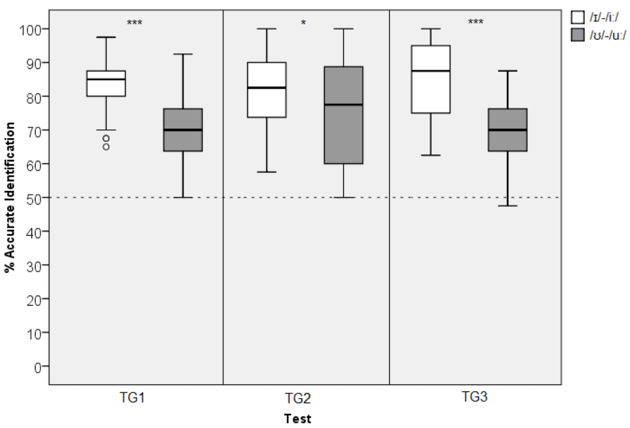


< .001] and Vowel [$F(1,34) = 24.19$, $p < .001$],indicating that the subjects improved their perception of both vowel pairs and the identification accuracy of the high-front vowel pair was consistently higher than the high-back pair. The interaction Test × Vowel was significant as well [$F(1,34) = 7.96$, $p = .008$]. Planned comparisons with Bonferroni correction on Test × Vowel interaction showed that the identification accuracy of both vowels pairs improved from pretest to posttest with significance (both at $p < .001$); and the high-front pair was significantly more accurately identified than the high-back pair in the posttest ($p < .001$), but not in the pretest ($p = .164$). In the posttest, the correct identification of the high-front vowel pair was 9.79% more than the high-back pair.

3.1.2. Tests of Generalization

The boxplot in Figure 2 combines the results of all three Tests of Generalization.

Figure 2: Mean percentages of accurate identification of the two vowel pairs (high-front = white boxes; high-back = dark boxes) in the all three TGs [*** $p < .001$; * = $p < .05$]. The dashed line indicates the chance level performance.

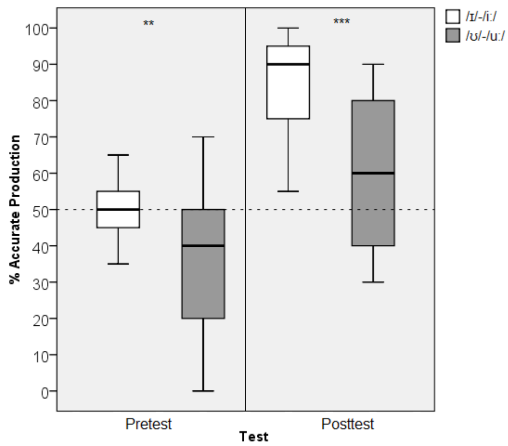


A two-factor ANOVA with TG (TG1, TG2, TG3) and Vowel (high-front, high-back) as factors was conducted to compare the transfer of perceptual learning to new speakers and/or new words. The main effect of Vowel [$F(2,68) = 59.49$, $p < .001$] was significant and so was the interaction Test × Vowel [$F(2,68) = 7.35$, $p = .001$]. The high-front vowel pair was again more accurately perceived than the high-back vowel pair in new speaker and word contexts. Planned comparisons with Bonferroni correction on Test × Vowel interaction demonstrated that the high-front vowel pair was significantly more accurately identified than the high-back vowel in all three TGs (TG1: $p < .001$; TG2: $p = .026$; TG3: $p < .001$). Yet, the main effect of TG was not significant ($p = .407$), indicating that the subjects' identification scores when listening to new words and new speakers were as accurate as with familiar words and familiar speakers.

3.2 Transfer of Perceptual Learning to Production

Figure 3 displays the results of production pretest versus posttest of the two vowel pairs:

Figure 3: Mean percentages of target production of the two vowel pairs (high-front = white boxes; high-back = dark boxes) in the pretest (left) and the posttest (right) [*** $p < .001$; ** $p = .001$]. The dashed line indicates the chance level performance.



A two-way repeated-measures with Test (pretest, posttest) and Vowel (high-front, high-back) as factors showed significant main effects of Test [$F(1,34) = 121.83$, $p < .001$] and Vowel [$F(1,34) = 30.81$, $p < .001$] since all the subjects improved their production of the two vowel pairs from pretest to posttest whereas the high-front vowel pair was more accurately produced than the high-back pair (mean difference = 17.29%). The interaction Test × Vowel [$F(1,34) = 6.19$, $p = .018$] was also robust. Planned comparisons with Bonferroni correction on Test × Vowel interaction showed that both production accuracy from pretest to posttest improved significantly for both vowel pairs (both at $p < .001$) more target productions were found for the high-front vowel pair than the high-back pair in both the pretest ($p = .002$) and the posttest ($p < .001$), indicating that the high-back vowel remained more difficult to be accurately produced than the high-front vowel pair.

3.3 Acoustic Analysis on Production Data

F1-F2 spaces in Figure 4 compare and contrast the production of the vowel pairs before and after training (due to page limit only data from female subjects is shown). In the pretest, the subjects produced /ɪ/ like /i:/ and /ʊ/ like /u:/. After training, the vowels produced were more separated, showing that the subjects started to produce the four vowels with different vowel qualities.

Figure 4: F1-F2 spaces of the two target vowel pairs, /ɪ/-/i:/ and /ʊ/-/u:/ before and after training.

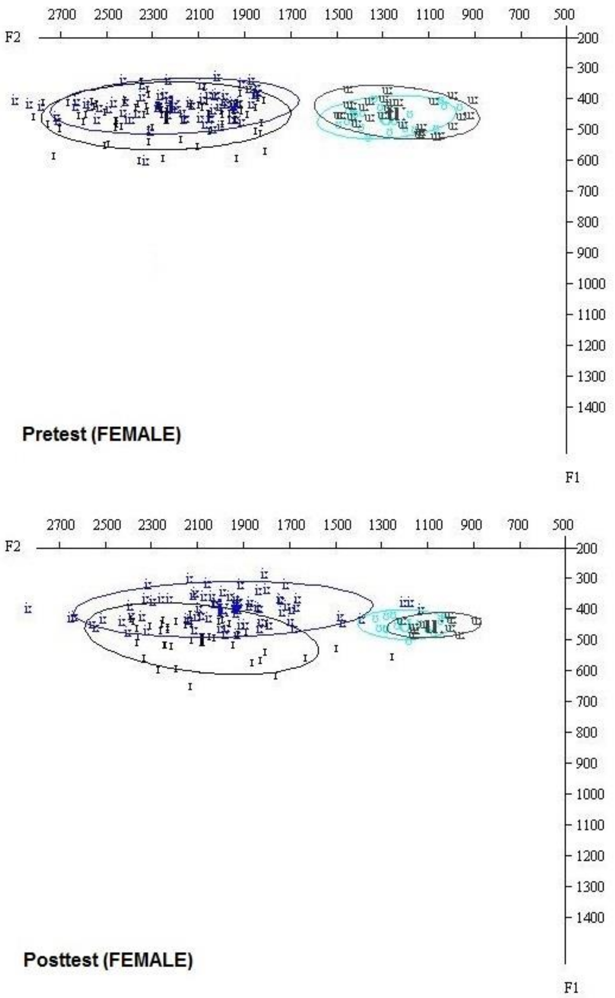


Table 1 summarizes the vowel duration of the two vowel pairs before and after training by both male and female subjects. All the subjects consistently produce the two long vowels with longer vowel durations.

Table 1: Average vowel duration (measured in milliseconds) by the subjects before and after training (standard deviations in brackets)

MALE	/ɪ/	/i:/	/ʊ/	/u:/
Pretest	148 (25)	193 (68)	172 (35)	192 (40)
Posttest	131 (21)	221 (63)	185 (36)	253 (52)
FEMALE	/ɪ/	/i:/	/ʊ/	/u:/
Pretest	147 (24)	192 (35)	182 (33)	204 (49)
Posttest	130 (27)	247 (28)	209 (45)	249 (47)

4. DISCUSSION AND CONCLUSIONS

This study further confirmed previous reports showing the effectiveness of adopting multi-speaker highly-variable stimuli to allow learners to attend to specific and relevant acoustic cues necessary to develop more language-specific phonetic categories.

Despite the general success of HVPT in training the two vowel pairs, one interesting finding is that the high-front vowel pair /ɪ/-/i:/ was consistently better identified and produced than the high-back vowel pair /ʊ/-/u:/ and more individual differences were observed in the high-back vowel. Acoustic analysis also showed that the F1 and F2 values of the back vowel pairs produced by the subjects, though having begun to show signs of separation (signifying that training helped improve the production in terms of F1 and F2), still remained more overlapped than the high-front vowel pair even after training. The result is not totally in accord with SLM which predicts that L2 phonemes which are closer to an L1 category will be more difficult to learn because the present study shows that even there are vowels in Cantonese which are similar to the target ones, the degree of learning of the two English contrasts can still be different. This may be because the back vowel pair is inherently harder to learn or due to the extremely scarce minimal instances that can be found. It might also be that the HVPT simply cannot provide as much learning effect as to the high-front vowel pair. Further research is demanded.

This study also showed that the subjects improved not just in terms of vowel duration (Bohn [2] found that it is universal for learners to take advantage of durational differences to discern contrasts), but also vowel qualities in the production. Cantonese speakers were found to be affected more by vowel quality cues than durational cues when they perceived their native vowels [18], future research can investigate what cues these subjects rely on when learning L2 vowels, so that future training paradigms can incorporate stimuli with cue manipulations to raise the awareness of the learners.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2005. *PRAAT: Doing phonetics by computer.* Retrieved from http://www.fon.hum.uva.nl/praat/.

[2] Bohn, O.-S. 1995. Cross-language speech perception in adults: L1 transfer doesn't tell it all. In Strange W., (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Timonium, MD: York Press, 275-300.

[3] Bradlow, A., Pisoni, D., Akahane-Yamada, R., Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.*, 101, 2299-2310.

[4] Chan, A. Y. W., Li, D. C. S. 2000. English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems. *Lang. Cul. and Curri.*, *13*, 67-85.

[5] Flege, J. E. 1995. Second-language speech learning: Theory, findings and problems. In Strange W., (ed.), *Speech perception and linguistics experience: Issues in cross-language research*. Timonium, MD: York Press, 233-272.

[6] Hung, T. T. N. 2005. Towards a Phonology of Hong Kong. In: Bolton K., (ed.), *Hong Kong English: Autonomy and Creativity*. Hong Kong: Hong Kong University Press, 119-140.

[7] Iverson, P; Hazan, V; Bannister, K. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *J. Acoust. Soc. Am.*, 118(5), 3267 - 3278.

[8] Iverson, P., Evans, B. G. 2009. Learning English vowels with different first language vowel systems II: Auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am.*, 126, 866-877.

[9] Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C. Molholt, G. 2005. The effects of identification training on the identification and production of English vowels by native speakers of Japanese. *Appl. Psycholing.*, 26, 227-247.

[10] Meng, H., Zee, E., Lee W. S. 2007. A Contrastive Phonetic Study between Cantonese and English to Predict Salient Mispronunciations by Cantonese Learners of English. Unpublished article. The Chinese University of Hong Kong.

[11] Nishi, K., Kewley-Port, D. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets. *J. of Speech, Lang., and Hear. Research*, 50, 1496-1509.

[12] Nishi, K., Kewley-Port, D. 2008. Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *J. of Speech, Lang., and Hear. Research*, 56, 1480-1493.

[13] Strange, W., Dittmann, S. 1984. Effects of discrimination training on the perception of /r/-/l/ by Japanese adults learning English. *Percept. and Psychophy.*, 36, 131-145.

[14] Wang, X. 2002. *Training Mandarin and Cantonese Speakers to identify English vowel contrasts: long-term retention and effects on production*. Unpublished doctoral dissertation, Simon Fraser University.

[15] Wang, X., Murray, J. M. 1999. The perception of English tense-lax vowel by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. *Proc. 14th ICPhS* San Francisco.

[16] Wong, J. W. S. 2012. Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. *Proc. SST2012* Sydney, 37-40.

[17] Wong, J. W. S. 2013. The Effects of Training Diversity in Training the Perception and Production of English Vowels /ɪ/ and /i:/ by Cantonese ESL learners." *Proc. Interspeech2013* Lyon, 2113-2117.

[18] Zhang, C. C., Peng, G., Wang, W. S.-Y. 2011. Effect of language experience on the categorical perception of Cantonese vowel duration. *Proc. Interspeech2011* Florence, 169-172.