THE EFFECTS OF HIGH-VARIABILITY PHONETIC TRAINING ON CANTONESE ESL LEARNERS' PRODUCTION OF ENGLISH /ɪ/-/iː/ CONTRAST - AN ACOUSTIC ANALYSIS

Janice Wing Sze Wong

Hong Kong Baptist University
janicewong@hkbu.edu.hk

ABSTRACT

This study investigated the effects of High Variability Phonetic Training (HVPT), a perception-only training, on Cantonese ESL learners' production of English /ɪ/-/iː/ contrast in terms of the first and second formant frequencies as well as the vowel duration. A total of 17 subjects participated in this study and they were all trained over 20 sessions of the HVPT. The results showed that after training, the male learners could improve their production of the two vowels in terms of F1, F2 and vowel duration; the female subjects only improved in terms of F1 and vowel duration. This suggested that perceptual learning can be transferred to the production domain and the subjects could generally improve their productions in terms of both vowel height and vowel length. More individual variations were observed in the change of vowel frontness/backness. This study showed that after training, the awareness of the differences, particularly for the length, between the target vowel contrasts was raised.

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

1. INTRODUCTION

A number of studies (e.g. [3], [10]-[11], [13], [18]-[19], [21]-[23]) have shown that the perception and production of non-native contrasts could be improved after phonetic training, one of which was the High Variability Phonetic Training (HVPT). HVPT is a perception-only phonetic training approach which emphasizes the use of multiple speakers and various phonetic contexts to increase the stimulus variability of the natural training minimal pairs. Subjects under training can be exposed to the natural tokens produced by different speakers. This approach was firstly adopted in a series of research studies ([3], [14]-[16]) in investigating the effectiveness of stimulus variability on the perceptual training of /ɹ/ and /l/ contrast of Japanese speakers of English. In the last two decades, this approach has been extensively researched and many displayed encouraging results in improving the subjects' perceptual performance of other confusing non-native contrasts and most of them showed that even though only training in the perception domain was given, the production of the contrasts could be improved. However, many of these studies only reported the production performance in terms of native speakers' judgement, without delving into the acoustic dimension to see what aspects of the vowels produced by the subjects have changed.

The present study has chosen to investigate the production learning of English vowels /ɪ/ and /iː/ by Hong Kong Cantonese speakers of English. This pair of vowels has been reported as posing difficulties in the L2 learners' production [5], [9], [17]. Even though most of the learners can distinguish the two vowels in terms of the vowel length, they are not aware that the two vowels differ also in terms of vowel quality. A couple of HVPT studies (e.g. [21], [23]) reported positive results in the modification of /ɪ/-/iː/ production by Hong Kong Cantonese speakers after training, but the acoustic properties of the production data were seldom analysed. This study aimed to complement the understanding of the effectiveness of the HVPT paradigm as well as the product of learning through a perception-only training paradigm.

2. METHODOLOGY

2.1. Participants

A total of 17 native Hong Kong Cantonese-speaking secondary school students (9 females and 8 males) aged around 16 to 17 were recruited to participate in the current experiment. They completed production pretest and posttest as well as 20 perceptual training sessions. They had not resided in any English-speaking countries before and the average age of learning English as an L2 was 3.71 (SD = .36), an average of 12.70 years (SD = .34). They reported no history of hearing or speaking impairment.

For the perceptual training and tests, another 6 native Received Pronunciation English speakers (3 females and 3 males) were recruited to produce the perceptual stimuli. Their ages ranged from 20 to 45.

2.2. Production Test Design

The aim of this study was to test how the HVPT paradigm can affect the production of the front vowel non-native contrast /ɪ/-/iː/ in terms of F1 and F2 values as well the vowel duration. The same production test was given to all the subjects before and after the 20 sessions of HVPT training.

All the subjects were given a word list with 20 words with the target vowels (10 /ɪ/ and 10 /iː/) plus 10 distractors. These target words were only a subset of the perceptual training set. 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). The instructions for this production pretest were offered to the subjects in the form of five practice trials and they had to produce them with natural loudness and speaking rate. They were not provided with any audio prompts or instructions before or during the recording. They could also pause and resume during the recording based on their own pace. The test took around 15 minutes to complete.

2.3. Stimulus Materials

Six native Received Pronunciation English speakers produced all the stimuli used in the HVPT training. All of them produced all the minimal word pairs used in the training. All words were CVC monosyllabic words with different onsets and codas.

Each stimulus word was produced three times to avoid intra-speaker variability. All the apparatus and setting used in the preparation of perceptual materials were the same as those used in the production tests.

2.4. Perceptual Training Design

A total of 40 stimuli (20 /ɪ/ and 20 /iː/) were presented to the subjects. All the stimuli were produced by six different native speakers to enhance the stimulus variability. The words were randomized in terms of speakers and word order in each session. The subjects were trained on a two-alternative forced choice paradigm. The stimuli were one of the counterparts in a minimal word pair contrasting the two vowels, e.g. among "bid" and "bead," only one of them was chosen for the test item. During training, immediate feedback was given; at the end of each session, their total scores were also shown. Each session required around 10 minutes to complete. All the subjects participated in two sessions each day for 10 consecutive days.

2.5. Method of Analysis

All the production data in the pretest and posttest was analysed acoustically by using the *Praat* speech analysis software [1]. The duration of the vowels was measured whereas the first two formant frequencies were gauged at the midpoint to evaluate how similar or different the vowel productions were before and after training.

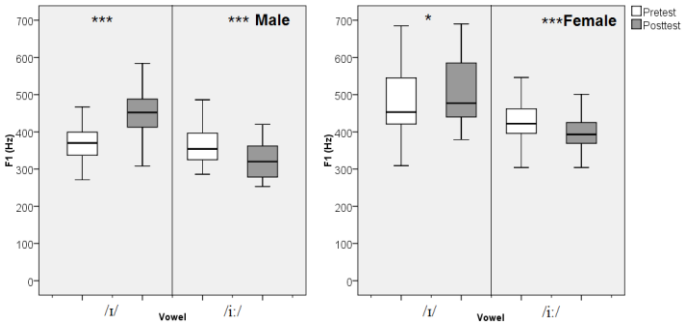
In total, there were 680 tokens (17 subjects × 20 words × 2 tests) produced by all 17 subjects. Yet, 23 tokens were not analysed since the subjects produced them as other words. Tokens produced by male and female subjects were analysed separately.

3. RESULTS

3.1. Vowel height: F1 values

Figure 1 juxtaposes the F1 values in English /ɪ/ and /iː/ productions by both male and female subjects at pretest and posttest.

Figure 1: Boxplots of F1 in English /ɪ/ and /iː/ productions by both male (left) and female (right) subjects at pretest and posttest



For male participants, a two-way repeated measures ANOVA with Test (pretest and posttest) and Vowel (/ɪ/ and /iː/) as factors revealed significant main effects of Test [$F(1,153) = 7.31$, $p = .01$] and Vowel [$F(1,153) = 52.7$, $p < .001$] due to the change of F1 values before and after the training. The interaction Test × Vowel was also significant [$F(1,153) = 70.24$, $p < .001$]. Planned comparisons revealed that the F1 of /ɪ/ increased by 77.77 Hz after training ($p < .001$) whereas that of /iː/ decreased by 41.30 Hz after training ($p < .001$). The F1 values of the two vowels were not significantly different in the pretest ($p = .519$). This result suggests that the male subjects have started to distinguish the two vowels and have learnt the vowel height differences between the two target vowels.

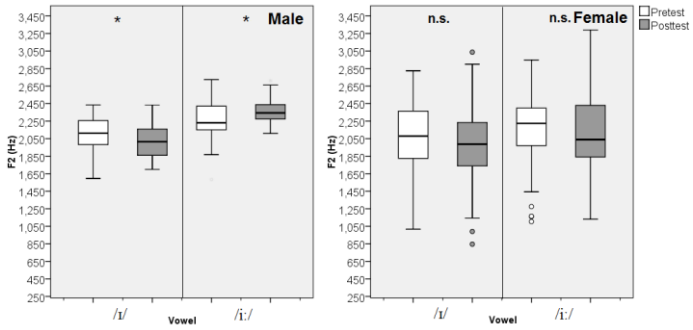
While for the female participants, only the main effect of Vowel was significant [$F(1,166) = 157.39$, $p < .001$]; the effect of Test was not ($p = .524$). Yet, the interaction Test × Vowel was significant [$F(1,166) =$

16.87, $p < .001$]. Planned comparisons showed that the F1 increased by 20.54 Hz ($p = .044$) for vowel English /ɪ/ after training whereas the F1 decreased by 27.82 Hz for /iː/ ($p < .001$). Yet, the female subjects could distinguish the two vowels at the pretest (F1 of /ɪ/ minus /iː/ = 61.14 Hz; $p < .001$); it was only the difference in F1 of the two vowels that was increased in the posttest (F1 of /ɪ/ minus /iː/ = 109.56 Hz; $p < .001$). This result shows that after training, the female subjects had more awareness of the difference of the two vowels in terms of vowel height and they tended to increase the vowel height differences between the two vowels.

3.2. Vowel frontness/backness: F2 values

Figure 2 shows the target vowel production performance in terms of F2 values by both male and female subjects at pretest and posttest.

Figure 2: Boxplots of F2 in English /ɪ/ and /iː/ productions by both male (left) and female (right) subjects at pretest and posttest



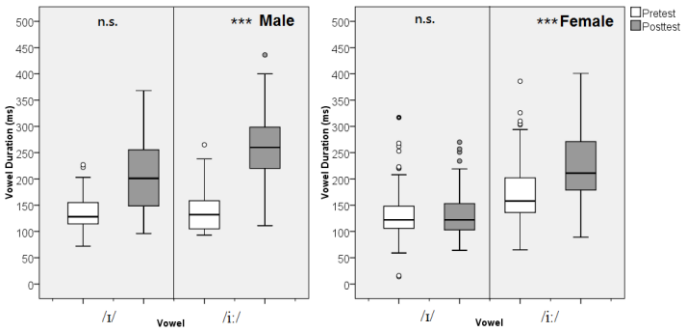
For male participants, a two-way repeated measures ANOVA with Test (pretest and posttest) and Vowel (/ɪ/ and /iː/) as factors revealed that there was only a significant main effect of Vowel [$F(1,153) = 85.09$, $p < .001$] but not for Test ($p = .822$). The interaction Test × Vowel was significant [$F(1,153) = 9.58$, $p = .003$] and planned comparisons showed that the F2 of /ɪ/ was decreased by 87.58 Hz after training ($p = .032$) whereas that of /iː/ was increased by 76.35 Hz after training ($p = .024$). However, in the pretest, the male subjects could already make a distinction between the two target vowels (F2 of /iː/ minus /ɪ/ = 183.07 Hz; $p < .001$). The difference in F2 of the two vowels (/iː/ minus /ɪ/) was increased to 346.98 Hz after training, suggesting that the male subjects were more aware of the differences of the two vowels in terms of frontness and backness of the vowels.

While for female participants, the same ANOVA showed that only a significant main effect of Vowel was observed [$F(1,166) = 10.03$, $p = .002$] because the two vowels were well distinguished in terms of vowel frontness/backness by the subjects. However, that of Test was not robust ($p = .285$) and neither was the interaction Test × Vowel ($p = .577$). This result showed that even after training, the female subjects generally did not alter the F2 much in the production of the two vowels. It is also worth noting that more individual variation was observed among the female subjects than the male ones.

3.3. Vowel duration

Figure 3 displays the vowel duration values of English /ɪ/ and /iː/ productions by both male and female subjects at pretest and posttest.

Figure 3: Boxplots of vowel duration in English /ɪ/ and /iː/ productions by both male (left) and female (right) subjects at pretest and posttest



Both male and females subjects performed similarly. A two-way repeated measures ANOVA with Test (pretest and posttest) and Vowel (/ɪ/ and /iː/) as factors revealed significant main effects of Test [$F(1,153) = 15.69$, $p < .001$] and Vowel [$F(1,153) = 166.14$, $p < .001$] for the male subjects. The interaction Test × Vowel was also robust [$F(1,153) = 15.94$, $p < .001$], due to the fact that the subjects produced a longer vowel /iː/ in the posttest than in the pretest (an increase of 55.45 ms, $p < .001$), although the subjects already distinguished the two vowels in terms of vowel length in the pretest.

The same ANOVA was conducted on the data of the female subjects. The main effects of Test [$F(1,166) = 24.27$, $p < .001$] and Vowel [$F(1,166) = 175.55$, $p < .001$] were robust and so did the interaction Test × Vowel [$F(1,166) = 25.87$, $p < .001$]. Similarly, the female subjects lengthened the vowel /iː/ in the posttest (an increase of 50.31 ms, $p < .001$). The result showed that all the subjects tended to exaggerate the vowel length of the long vowel /iː/ and after training they were more aware of the differences in terms of vowel length between the two target vowels.

4. DISCUSSION

This study has complemented previous training studies by showing what aspects of the English/ɪ/ and /iː/ productions the subjects had changed after HVPT training. The results display that after 20 sessions of HVPT, both male and female subjects could improve their production of both /ɪ/ and /iː/. Concrete learning in production can be observed as the acoustic values have changed and more distinction was found between the two vowels. From previous studies, the two English vowels were found to be overlapped in terms of F1 and F2 since this contrast does not exist in Cantonese phonological system. Only a similar pair (/i/ and its surface form [ɪ] with overlapping acoustic properties) exists in Cantonese. Thus, the pretest result is in accord with SLM [8] as L2 sounds which are similar to the L1 system are more difficult to acquire. The promising results of the present study then show that the HVPT is effective in modifying a difficult contrast. Highly variable stimuli are beneficial in leading to a significant level of success in non-native contrast acquisition as the stimuli could promote selective attention of the subjects. Offering a wider range of stimuli to the subjects is also believed to be able to expand the exemplar space along the dimensions where the two vowels differ and shrink along the dimensions that do not show distinctions for the two vowels, thus leading to more accurate productions.

The results also show that the awareness of the subjects towards vowel height and vowel length was raised, whereas that of vowel frontness/backness show inconclusive results because the female subjects did not demonstrate significant changes in the production of the two vowels in terms of F2. Although this study did not aim to investigate what cues the subjects relied on, the present results still suggest that the subjects had more awareness and reliance on vowel height and vowel length as cues to help distinguish the two vowels. The subjects even exaggerated the vowel length of the long vowel in the posttest although they could distinguish the pair in terms of vowel duration in the pretest. This result hints that the subjects tended to rely heavily on duration as the cues to distinguish a contrast, similar to previous reports (e.g. [2], [4], [6], [7], [12]). However, Zhang, Peng and Wang [24] discovered that Cantonese speakers were affected more by vowel quality cues than durational cues when they perceived Cantonese vowels. After training through the HVPT paradigm, the subjects also improved the ability to distinguish between the differences of the two vowels in terms of vowel height. This piece of result might lend some support to the above study that duration alone may not be an efficient cue in categorization by Cantonese speakers in both L1 and L2; rather, vowel quality cues overrode the effect of duration. Further experiments particularly on the cues that this group of speakers rely on are demanded. This will reveal more about the link between perception and production as well.

Useful pedagogical information was obtained from the positive results in the study which indicated that there exists the possibility of successfully training ESL learners to produce a non-native contrast more accurately. While maintaining the viability of application and simplicity of procedures of the training was also of high importance for language teachers, HVPT is practical enough to be adopted easily in schools or learning centers since it is not a complicated design. This paradigm can even be setup online for learners who are willing to receive training at their own pace and time. It will be worth investigating the optimal number of training sessions and training intensity which can successfully promote successful learning so that the training can be adopted in L2 classrooms and bring the most benefits to both learners and teachers.

Future studies can further assess the learning of the subjects by adopting and comparing different training types such as production training, audio-visual training etc. The adoption of assessment tasks such as identification test, category discrimination test, cue weighting test, etc. can also be investigated as they all tap into different aspects of L2 vowel perception and production. More efforts should also be directed at individual variability among the subjects both before and after training. The improvement of the subjects shown in the present study only gave an evaluation of the HVPT paradigm at the aggregate level without considering individual performances. From this study, there are clearly some learners who benefited from training more than others. Some previous studies (e.g. [20]) have also shown that individual's performance at the pretest is a good predictor of learning outcomes, and thus more efficacious training paradigms can be developed and variability found in different training studies can be eliminated when individual differences of subjects are taken into account.

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] Cebrian, J. 2002. Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast. *Toronto Working Papers in Linguistics Dissertation Series*. University of Toronto, Toronto, Canada.

[5] 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.

[6] Escudero, P. 2001: The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. In A. H.-J. Do, L. Dominguez & A. Johansen (Eds.). *Proc. 25th BUCLD* Cascadilla, 250-261.

[7] Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation, Ultrecht University).

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Kondaurova, M., Francis, A. L. 2004. Perception of the English tense/lax vowel contrast by native speakers of Russian. *J. Acoust. Soc. Am., 116*, 2572-2573.

[13] 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.

[14] Lively, S. E., Logan, J. S., & Pisoni, D. B. 1993. Training Japanese listeners to identify English/ɹ/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.*, 94, 1242-1255.

[15] Lively, S. E., Pisoni, D. B., Yamada, R. A., Tokhura, Y., & Yamada, T. 1994. Training Japanese listeners to identify English /ɹ/ and /l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am.*, 96, 2076-2087.

[16] Logan, J. S., Lively, S. E., & Pisoni, D. B. 1991. Training Japanese listeners to identify English /ɹ/ and /l/: A first report. *J. Acoust. Soc. Am.*, 89, 874-886.

[17] 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.

[18] 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.

[19] 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.

[20] Perrachione T. K., Lee, J., Ha L.Y., Wong P. C. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Jrl. of Acou. Soc. of America, 130*, 461-472.

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

[22] 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.

[23] 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.

[24] 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.