The Effects of Order and Intensity of Training on the Perception and Production of English /e/-/æ/ by Cantonese ESL learners

Janice Wing Sze Wong

Hong Kong Baptist University
janicewong@hkbu.edu.hk

Abstract

This study investigated the effects of the order of provision of perception and production training and the intensity of the two training types on the perception and production of English /e/-/æ/ by Cantonese ESL learners. Thirty-five subjects were assigned into four groups with different training order and intensity levels. Results showed that all groups improved their perception and production of /e/-/æ/, but the groups which received production training before perceptual training performed better than those who were trained perceptually first. Training intensity, i.e. how many days the training sessions were spread over, however, did not affect the degree of improvement.

Index Terms: High Variability Phonetic Training, explicit articulatory training, training order, training intensity

1. Introduction

Research studies conducted in the last few decades have shown that difficulties in learning L2 contrasts can be ameliorated by laboratory training such as perceptual-only training (e.g. [1]-[2]), production-only training (e.g. [3],[4]), or a combination of both (e.g. [5],[6]). These studies have shown inconsistent results although most of them reported some improvement of learning in either/both modalities. The present study aimed to evaluate the effectiveness of both perception and production training on the perception and production of the English /e/-/æ/ contrast among Hong Kong Cantonese ESL learners. This vowel pair is commonly-confused by this group of learners [7]. Also, it would be useful in practice to examine the effects of training order, i.e. training in one modality prior to the other and the reverse, as previous research (e.g. [5],[6]) had not regarded this as a variable.

Meanwhile, training intensity had also been assumed as a constant in training studies and the consensus was that intensive training could contribute to the learning of these difficult contrasts. Thus, how intensive a treatment should be so as to achieve optimal results, or whether different intensity levels could generate diverse training effects, have been overlooked. No training studies on L2 speech perception and production have systematically investigated the effect of training intensity on subjects' performance; only in recent decades had training intensity begun to receive some attention in research on young children with language delays and disabilities and their results were not consistent (e.g. [8],[9]).

Thus, besides examining the effects of training in both modalities and the order of provision of training on the perception and production of /e/-/æ/ contrast, the effects of training intensity were also addressed in this study in order to garner some insights in ways to optimize training effects and provide pedagogical insights to teachers and learners.

2. Methodology

2.1. Design

All the participants took part in these three phases:

PHASE 1. Pretest Phase: including one production pretest and one perception pretest;

PHASE 2. Treatment Phase: depending on the training intensity (standard, one session per day for 20 days vs. intensive, five sessions per day for four days) and order (entire HVPT paradigm completed first followed by production training, vs. the reverse)

a. *High Variability Phonetic Training*: 10 training sessions were offered, each lasted for 10 minutes.

b. *Explicit Articulation Training*: 10 training sessions were offered, each lasted for 10 minutes.

PHASE 3. Posttest Phase: including one production posttest, one perception post-test, and three perception Tests of Generalization (TG1, TG2 and TG3).

2.2. Participants

A total of 35 secondary school students were recruited to the present experiment. They had Hong Kong Cantonese as their L1 and English as the L2 and were aged around 16 to 17. They all started learning English as an L2 at the age of 3.32 (SD = .53) for an average of 13.6 years (SD = .82). No one had resided in any English-speaking countries. They all reported that they had no hearing or speaking impairment.

All the subjects received the same number of training sessions and finished training in one modality first before another. They were hence divided into four groups depending on the training order and intensity:

- Group 1 – HPS: 9 received 10 **H**VPT perception training sessions, followed by 10 explicit articulatory **p**roduction training sessions and they completed one session per day for 20 days (**s**tandard training);
- Group 2 – HPI: 9 received also **H**VPT first before the **p**roduction training, but they received five training sessions per day for four days (**i**ntensive training);
- Group 3 – PHS: 9 subjects received the **p**roduction training first before **H**VPT and they received one session per day for 20 days (**s**tandard training);
- Group 4 – PHI: 8 received also the **p**roduction training first before **H**VPT but with five training sessions per day for four days (**i**ntensive training).

A total of eight native General American (GA) English speakers (6 female and 2 male) were also invited to produce stimuli for the tests and training. Their ages ranged from 35 to

40.

2.3. Setting and Apparatus

All the subjects completed the tests and training sessions in a language laboratory. They completed all perceptual training and test sessions by using a computer program designed by the researcher. They listened to the audio tokens presented in the program and completed an identification task. All the data were saved into a Microsoft Access database for analysis.

In the explicit production training sessions, videos in which a native GA speaker displayed and taught the articulation of vowels were given to the subjects to watch before practicing the target vowels with the researcher. The subjects recorded the production test tokens using Adobe Audition 1.5 through a Shure SM58 microphone.

2.4. HVPT Stimuli

Six of the eight native GA English speakers produced stimuli used in both the perceptual pre/posttest and the HVPT training. All of them produced 20 minimal word pairs, i.e. a total of 40 tokens. All words were CVC monosyllabic words with different onsets and codas. One of the speakers, i.e. a familiar speaker to the subjects, recorded also a new word list with 20 /e/-/æ/ minimal word pairs for the use in TG2 (new words by a familiar speaker). Another speaker whose voice had not appeared in the training or tests recorded the same set of tokens for the use in TG3 (familiar words by a new speaker). The last speaker who had not recorded any tokens for the training stimuli or the tests, i.e. a new speaker, recorded another new list with 20 /e/-/æ/ minimal word pairs for the use in TG1 (new words by a new speaker). All the minimal pairs in TGs were with various CVC contexts and syllable structures (mono-, di- and poly-syllabic) with a view to testing the transfer of learning under various conditions.

2.5. Production Training Materials

A female native GA speaker whose voice had not appeared in any perceptual stimuli recorded the training items. The video recordings were made in a soundproof room and the face of the speaker was put against a blue background. A Canon EOS 600D digital camera with video recording function was used. Full HD 1080p video recording at 25 frames were made. Audio was recorded using Adobe Audition 1.5 through reading into a Shure SM58 microphone with the X2u XLR-to-USB adapter for digital audio recording and was synced to the video afterwards.

2.6. Procedure

2.6.1. Pretest Phase

All groups participated in both the production and perception tests in the first phase. The production pretest was administered first to avoid subjects' cueing or being exposed to the items which would appear later in the perception pretest.

- *Production Pretest:* The subjects were given a word list of 20 words (with 10 /e/ and 10 /æ/) and had to record all the words which would appear either in the perception pretest or the training. To ensure authentic performance and that the subjects could produce also other segments apart from the vowel, before the pretest, the subjects could hear the pronunciations of the words produced by a native speaker who had not been involved in the study. 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 during the recording. They could also pause and resume during the recording based on their own pace. The test took less than five minutes to complete.
- *Perception Pretest:* The subjects could get access to the computer program to complete the identification test which included 50 questions (40 words with either /e/ or /æ/ with 10 distractors). They did 10 practice trials before the test, which were not analyzed. Each stimulus could be played by the subjects as many times as they needed before they chose the answer from three choices with conventional English orthography, or a blank for a free answer, in which they could type their own word. The frequency of occurrence of the correct answer that appeared in the four serial positions, i.e., word 1, word 2, word 3, free answer, were equal; thus the chance level was 25%. This design was an attempt to avoid using simply two choices with 50% of chance level. The program was also designed to limit the subjects from not answering one question before moving onto another. The whole test could be completed within 10 minutes.

2.6.2. Treatment Phase

Both HPS and HPI groups participated in HVPT first before the production training whereas PHS and PHI groups received the production training before HVPT. Details about the two types of training in this phase were as follows:

- *High Variability Phonetic Training:* A total of 40 stimuli (20 /e/ and 20 /æ/) produced by six different native English speakers, all randomized in terms of speakers and word order in each session, were presented to the subjects. The subjects were trained on a two-alternative forced choice identification task which directed their attention to identifying the target word and raised the training effect, unlike the four choices used in the pretest. During training, immediate feedback was given; at the end of each session, their total scores were also shown.
- *Explicit Production Training:* The subjects were provided with videos in which a native speaker demonstrated the articulation of the vowels. The subjects had to watch the video first and read after the video host. A word list with 20 different words containing one of two target vowels was given to the subjects and they pronounced after the researcher at least three times and immediate corrective feedback was given. Articulation information of the vowel pairs, i.e. the tongue position, vowel openness, as well as the length of the vowels was also emphasized in the each session explicitly with pictures as illustrations and mirrors to help them better produce the target sounds.

The "I" groups and "S" groups also differed in terms of the number of training sessions received per day. The "I" groups received five training sessions per day and in between each session, a short break with refreshments was given; the "S" groups received one training session per day. Thus, the "I" groups would complete all the training within 4 days whereas the "S" groups spent 20 days to finish the training.

2.6.3. Posttest Phase

The Posttest Phase involved one production posttest (same as

the Production Pretest) which was completed before the four perception posttests (posttest, TG1, TG2 and TG3), which were all done on the same day.

- *Production Posttest:* same as the Production Pretest
- *Perception Posttest:* same as the Perception Pretest
- *Test of Generalization 1:* The subjects heard 40 tokens (with 20 /e/ and 20 /æ/) spoken by a new speaker whose voice was not heard in any of the training stimuli or the tests. The procedures were similar to those administered in the Perception Pretest, and subjects were also given four choices to choose from.
- *Test of Generalization 2:* The subjects had to listen to 40 new words (with 20 /e/ and 20 /æ/) spoken by a familiar speaker, who had been one of the speakers in the training stimuli. Procedures were the same as those in TG1.
- *Test of Generalization 3:* The subjects revisited 40 familiar words which they had come across in the perception training sessions, but the words were produced by a new speaker. Again, the procedures were the same as those in TG1 and TG2.

2.7. Evaluation of Production Data

The production scores were evaluated by directly counting the number of accurate productions. The productions of the subjects were transcribed twice by a phonetically-trained researcher for whom Cantonese was the L1 and English the L2. The intra-rater reliability obtained was 92.30% (α = .833). Another researcher who had English as L1 also transcribed the data phonetically. During transcription, they transcribed phonetically the word they heard, which was not limited to only the target vowels. The reliability check was done without referring to any completed transcriptions. The inter-rater reliability was 91.36% (α = .815). A follow-up acoustic analysis on half of the productions, by checking the F1 and F2, F3 values and the vowel durations, was conducted by a third phonetically-trained researcher to confirm that the transcriptions aligned with the acoustic measures and were reliable. The acoustic analysis results were consistent with the transcription.

3. Results

3.1. Perceptual Performance

3.1.1. Effects of training: Pretest vs. Posttest

The following boxplot shows the results of the four groups in the pretest and posttest:

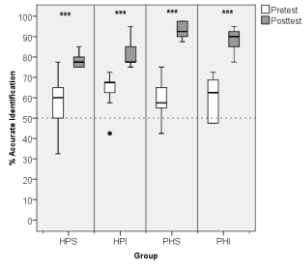

Figure 1: *Mean percentages of correct identification of the four groups in pre (white) & posttest (dark) [*** = p < .001].*

A four-way repeated measures ANOVA was computed using Test (pretest, posttest), Training Order (HP vs PH), Training Intensity (Standard vs. Intensive) and Vowel (/e/, /æ/) as factors. It showed significant main effects of Test [$F(1,31) = 182.94$, $p < .001$], Training Order [$F(1,31) = 5.30$, $p = .028$] and Vowel [$F(1,31) = 5.95$, $p = .021$]; yet, Training Intensity was not a significant factor ($p = .618$). The interactions Test × Order [$F(1,31) = 9.10$, $p = .005$], Test × Training Intensity [$F(1,31) = 7.44$, $p = .008$], Vowel × Training Intensity [$F(1,31) = 9.95$, $p = .004$] and Test × Vowel × Training Intensity [$F(1,31) = 4.93$, $p = .034$] were all significant. Planned comparisons with Bonferroni correction on Test × Training Order interaction showed that all groups improved significantly from pretest to posttest (both at $p < 0.001$). A significant difference between Training Order in the posttest, but not in the pretest ($p = 1.00$), was also found. In the posttest, PH groups outperformed HP groups by 10.61% ($p < .001$).

3.1.2. Generalizability of training

The following three boxplots show the results in TG1, TG2 and TG 3 (from left to right):

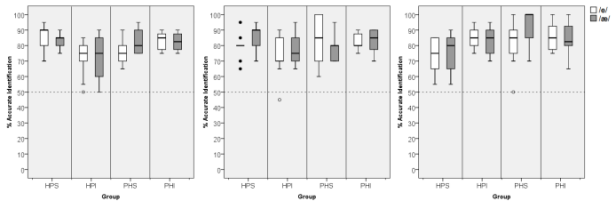

Figure 2: *Mean percentages of correct identification of vowels in all TG1, TG2 and TG3 (from left to right) across groups.*

For TG1 (new words produced by a new speaker), a three-way ANOVA with Training Intensity, Training Order and Vowel was computed. The result showed no significant main effect, but only one interaction, Training Intensity × Training Order, was robust [$F(1,31) = 10.80$, $p = .003$]. This was due to the fact that subjects trained under the PH order performed significantly better than the HP order in both intensity levels. A simple test of effects showed that only under intensive training condition did those subjects who were trained under PH perform better than HP for 11.15% ($p = .003$).

While for TG2 (new words produced by a familiar speaker), another three-way ANOVA showed that no main effect and interaction were significant. Yet, the figures still showed that the subjects could identify 79.43% and 81% accurately for the vowels /e/ and /æ/ respectively, meaning that the groups performed similarly.

With regard to TG3 (familiar words produced by a new speaker), the same ANOVA showed only a significant main effect of Training Order [$F(1,31) = 5.22$, $p = .029$] since the PH groups outperformed HP groups by 6.53%

Perceptual learning was shown to be able to generalize to new speakers and new tokens, and providing production training first before HVPT appeared to be more useful.

3.2. Production Performance: Effects of training (Pretest vs. Posttest)

A four-way repeated measures ANOVA was computed using Test (pretest, posttest), Training Order (HP vs PH), Training Intensity (Standard vs. Intensive) and Vowel (/e/, /æ/) as factors. Only the main effects of Test [$F(1,31) = 275.45$, $p < .001$] and Training Order [$F(1,31) = 4.80$, $p = .036$] were robust, indicating that all groups showed improvements in production accuracy after training and the order of training played a role in the learning. The interaction Test × Training Order [$F(1,31) = 11.58$, $p = .002$] was also significant.

Planned comparisons with Bonferroni correction showed that those subjects who were trained under the production training before HVPT outperformed the other two groups by 11.49% in terms of production accuracy ($p < .001$) while in the pretest their performance were similar ($p = .511$). However, neither the main effects of Vowel ($p = .531$) and Training Intensity ($p = .265$) nor the other interaction effects, were significant. This boxplot displays the results of production pretest versus posttest across groups:

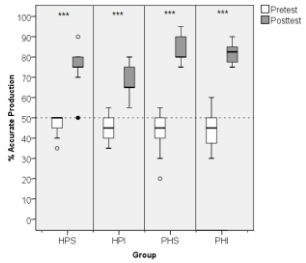

Figure 3: *Percentage of target production of the four groups between the pretest (white) & posttest (dark) [*** = p < .001].*

4. Discussion

The present results indicated that providing both high-variability phonetic training and explicit articulatory training could benefit L2 leaners in the perception and production of non-native vowel contrasts /e/-/æ/. The success could be attributed to the use of high variability stimuli in HVPT and the corrective feedback given in the production training. It is not clear, however, which training or exactly what elements in the training benefited the participants more, although this is not the goal of the present research. Still, it is found that the order of training plays a rather important role in determining the degree of success.

Training in production training before perception training helped the participants improve more in both the perception and production of the target vowel pair than the reverse order of training. This finding is intriguing as previous studies utilizing training in both modalities have not considered this as a factor. Modifying the articulatory patterns in production training first before being exposed to a wide variety of acoustic cues from the perceptual stimuli was more helpful in both perception and production. This appears to suggest that there exist some links between perception and production and perceptual representations may be heavily articulatorily-based [10]. Changes in the basic perceptual unit would lay an even more solid foundation for the tuning of perceptual representations as well as phonological-motor mapping when the learners were to receive further perceptual training later, than those who were trained in perception before production. Although the present study did not aim to test any theories, the findings suggested a crucial role played by articulatory gestures in the perception and production of non-native speech sounds. Yet, the present study has not traced the performance progress during training to gauge the amount of benefit brought by each of the two types of training or how training in one modality first can benefit the other. Further research that directly compares the effectiveness of perception-only and production-only on the two modalities would also be useful.

Another finding in this study was that training intensity did not show any significant effects in the learning in the two modalities. One plausible explanation was that the "active ingredients" of the training may matter more [11]. Intensity alone is insufficient to determine the training outcome; rather, it is the active ingredients (e.g. the type of training utilized, the training components used, how the training was delivered, etc.) that contribute to the learning. Following this line of reasoning, it is probably the adoption of high-variability perceptual stimuli, identification tasks, the use of corrective feedback in articulatory training, etc. that already become the contributing parameters leading to successful learning, lowering the possible effect that might be brought by training intensity. The optimal number of training that can improve learners' performance also merits more investigation.

5. Conclusions

The present study showed that training order had an effect on the perception and production of the English vowel pair /e/-/æ/ while training intensity did not. It suggested that when some training was provided, learning would occur and would not be affected by how the training sessions were spread over a period of time. It was rather the training order that influenced the training outcome. This finding may benefit second language learners who have difficulties in these non-native contrasts to train themselves at their own pace. Future research can investigate the effects of a wider variety of intensity levels and its interaction with other vowel training programs.

6. References

[1] Bradlow, A., Pisoni, D., Akahane-Yamada, R., and Tohkura, Y., "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, 1997.

[2] Iverson, P., and Evans, B.G., "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, 2009.

[3] Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N., "The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds," *J. Acoust. Soc. Am.*, 138(2):817-832, 2015.

[4] Saloranta, A., Tamminen, H., Alku, P., and Peltola, M. S., "Learning of a non-native vowel through instructed production training", *Proc. ICPhS2015*, Paper 0235, 2015.

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

[6] Aliaga-Garcıa, C., and Mora, J. C., "Assessing the effects of phonetic training on L2 sound perception and production," in M. A. Watkins, A. S. Rauber, and B. O. Baptista [Eds], *Recent Res. Sec. Lang. Phon./Phono.: Percep. Produ.*, 2-31, Cambridge Scholars Publishing 2009.

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

[8] Fey, M.E., Warren, S.F., Brady, N., Finestack, L.H., Bredin-Oja, S. L., Fairchild, M., Sokol, S., and Yoder, P. J., "Early effects of responsivity education/prelinguistic milieu teaching for children with developmental delays and their parents", *J. Speech, Lang., Hear. Res.*, 49:526-547, 2006.

[9] Gray, S., "Word-learning by preschoolers with specific language impairment: Effect of phonological or semantic cues," *J. Speech, Lang., Hear. Res.*, 48:1452-1467, 2005.

[10] Best, C. T., "A direct realist view of cross-language speech perception: New Directions in Research and Theory", in Winifred Strange [Ed], *Speech perception and linguistic experience: Theoretical and methodological issues*, 171-204, York Press, 1999.

[11] Baker, E., "Optimal intervention intensity", in *Inter. J. Speech-Lang. Path.*, 14:401-409, 2012.