Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training

Janice Wing Sze Wong

The Chinese University of Hong Kong

jwong_aletheia@cuhk.edu.hk

Abstract

This study compares the effectiveness of the High Variability Phonetic Training (HVPT) approach and the Low Variability Phonetic Training (LVPT) approach on training the perception and production of English /e/ and /æ/ by Cantonese ESL learners. In a pretest-treatment-posttest design, 22 subjects were trained under the HVPT and 19 under the LVPT. All subjects showed significant perceptual learning and generalization of the two vowels to new words and new speakers, with those trained under the HVPT demonstrating more improvement. Robust transfer of perceptual learning to production was also observed in both groups, with the HVPT approach yielding more improvement. Robust transfer of perceptual learning to production was also observed in both groups, with the HVPT approach yielding more improvement.

Index Terms: High Variability Phonetic Training, second language acquisition, speech perception and production

1. Introduction

It is typical to find second language (L2) learners having difficulties in acquiring novel phonetic categories. Linguists over the past few decades have started to examine the effects of laboratory-based phonetic training on the modification of perception and/or production of non-native contrasts, which are usually perceived and produced by L2 learners according to their L1 categories [1]. Among all, the High Variability Phonetic Training (HVPT) approach, which involves the use of natural training stimuli with various phonetic contexts produced by multiple speakers in an identification task with immediate feedback, has received particular attention due to its effectiveness in perceptual learning of different consonantal and vowel contrasts, such as in [2][3][4]. More research interest has also been drawn as the perceptual gain in the HVPT was found to also promote production improvement, suggesting an underlying link between perception and production [5].

In contrast to the Low Variability Phonetic Training (LVPT) approach that only adopts stimuli produced by one talker and in a single phonetic context (as in [6]), the HVPT is believed to enhance participants’ long-term recall and force them to focus on phonetic cues that underlie categorical distinctions. However, no research has been done to show how beneficial the HVPT is over the LVPT, i.e., how facilitative the stimulus variability is, by directly comparing the two approaches. The present study hopes to shed more light on this issue by investigating the effects of perceptual training on the perception and production of the English /e/-/æ/ contrast by a group of Hong Kong Cantonese ESL learners. This pair of contrast does not exist in Cantonese and has been found to pose a lot of difficulties to the learners in their perception and production [7]. Generalizability of the learning of this non-native contrast was also investigated.

2. Methodology

2.1. Participants

A total of 64 students, recruited from a secondary school in Hong Kong, participated in the experiment. They were divided into three groups: (1) 22 of them (12 female and 10 male) were trained under the HVPT approach; (2) 19 of them (11 female and 8 male) were trained under the LVPT; and (3) 23 of them were the control group (19 female and 4 male) receiving no training. The subjects, aged between 16 and 19 (average 16.97), were all local Hong Kong Chinese with Cantonese as their native language. They started learning English as an L2 in Hong Kong at 3.04 of age on average (for an average of 13.94 years). None had resided in any English-speaking countries. None reported hearing or speaking deficits.

Seven native English speakers (three female and four male) were also invited to produce stimuli for the tests and training. Their ages ranged from 35 to 60, with three speaking General American accent, one Inland Northern American, one North East English, and two South East English.

2.2. Design

The present study employed an experimental design in which two groups of subjects took part in either the HVPT or LVPT sessions for three weeks, with a control group receiving no treatment. All subjects participated in these three phases:

PHASE 1. Pretest Phase, which included one production pretest and one perception pretest;

PHASE 2. Perceptual Treatment Phase, with a total of 10 training sessions (either the HVPT or LVPT; except for the control group) in which they attended two sessions a day for three weeks; and,

PHASE 3. Posttest Phase, with one production posttest, one perception post-test, one production Test of Contextualization (TC) and two perception Tests of Generalization (TG1 and TG2).

2.3. Setting and Apparatus

All tests and training sessions took place in a language laboratory. A computer training program was designed by the researcher for all perceptual training and test sessions. The subjects listened to audio speech tokens presented by the program and data were saved into a Microsoft Access database. They recorded the production test tokens using Adobe Audition 1.5.

2.4. Stimulus Materials

The stimuli made by six of the seven native English speakers contributed to both the perceptual pre/posttest tokens and the...
perceptual training stimuli. For the HVPT stimuli, each speaker involved produced 10 pairs of /e/ and /æ/ real word pairs, contributing a total of 60 stimuli (6 speakers × 10 trials) with a wide variety of phonetic environments (with different syllable structures, onsets and codas). For the LVPT, one of the six speakers was invited to record all 60 word pairs. All the productions made by this speaker were also utilized in all the perception tests. Another one among the six native speakers, i.e., a familiar speaker to the subjects, also recorded a word list for TG2 (new words by a familiar speaker). The word list included 30 /e/ and /æ/ minimal word pairs which did not appear in the training sessions. The last speaker who had not recorded anything for the training stimuli or the pre/posttests, known as a new speaker, recorded another new list with 30 /e/ and /æ/ minimal pairs for TG1 (new words by a new speaker). All the minimal word pairs were with various CVC contexts and syllable structures.

Each speaker read the tokens at least three times, which avoided the use of a single token per speaker for the stimuli. All the stimuli were made by the native speakers reading into a headphone-mounted computer with Adobe Audition 1.5 software for digitization. The words were saved separately, digitized at a sampling rate of 128 Kbps with 44100 Hz, normalized for peak amplitude before incorporating them into the computer software.

2.5. Procedure

2.5.1. Pretest Phase

The first phase involved two tests: a production pretest and a perception pretest. All groups participated in both tests. 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 list of 60 words (30 /e/ and 30 /æ/) and 10 distractors. 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. In the test, the subjects had to record all the words that would appear either in the perception pretest or the training.

  The instructions of 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 at a natural speaking rate. They were not provided with any audio prompts or instructions during the recording. They could also pause and resume the recording at their own pace. The test took less than 10 minutes to complete.

- **Perception pretest**: The subjects used the computer program to complete this test within 30 minutes. There were 70 trials in total (60 words with either /e/ or /æ/ plus 10 distractors). They did five practice trials before the test, which were not analyzed.

  Before they confirmed their answers, they had to listen to the stimuli and choose 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 set was 25%. This four-alternative choice test design has not been adopted in any previous studies using the HVPT or the LVPT. This new design is more sensitive as it avoids using only 50% probability as chance level, as in common identification tests with only two choices.

2.5.2. Training Phase

The training was administered twice a day for three days in the first week and two days in the second week, making a total of 10 training sessions. In between the two training sessions, the subjects were given at least 30 minutes for rest.

- **The HVPT**: The subjects were presented with 60 stimuli produced by six different native English speakers, all randomized in terms of speakers and word order. They 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 “bed” and “bad,” only one of them was chosen for one test item). The subjects were deliberately provided with only two-alternative forced choices in all the training sessions so that they could pay more attention to and focus on identifying only the two vowels without interference from other alternative vowels. Unlike all test sessions in which four choices were given, the aim of having only two choices in the training sessions was to raise the intensity of the training effect. During training, immediate feedback was given; at the end of each session, their total scores were also shown.

- **The LVPT**: The training procedures were exactly the same as in the HVPT. The only difference was in the training stimuli: the same 60 words were produced by only one female native speaker of English.

2.5.3. Posttest Phase

The posttest phase involved two production posttests (posttest and TC) that were completed before the third perception posttests (posttest, TG1 and TG2).

- **Production posttest**: same as the production pretest.

  Test of Contextualization (TC): All subjects were given a 250-word passage which included 50 content words with /e/ and /æ/. They were asked to record the whole passage naturally, at their own pace and loudness.

- **Perception posttest**: same as the perception pretest

  Test of Generalization 1 (TG1): The subjects heard 30 new words 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 they were also given four choices, three provided choices and one open answer, to choose from.

  Test of Generalization 2 (TG2): The subjects had to listen to 30 new words spoken by a familiar speaker, who had been one of the speakers in the training stimuli. Procedures were the same as those in TG1.

2.6. Evaluation of Production Data

The production scores were evaluated by directly counting the number of accurate productions of the target vowels. 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 $r = .982$ ($p < .0001$). Another phonetically-trained Ph.D. student studying Applied English Linguistics with Cantonese as L1 and English as L2 also transcribed the data phonetically. The reliability check was done without referring to any completed transcriptions and it was 94.74%. The intra-rater reliability was $r = .907$ ($p < .0001$). A follow-up acoustic
analysis on half of the transcriptions, by checking the F1, F2, F3 values and the vowel durations, was also done to confirm if 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

Figure 1 above shows the overall results of the identification performance of the HVPT, LVPT, and control groups.

A three-factor repeated measures analysis of variance (ANOVA) was computed using group (HVPT, LVPT, control), test (pretest, posttest), and vowel (/ɛ/, /æ/) as factors. It showed significant main effects of group \( F(2,61) = 11.66, p < .0001 \), test \( F(1,61) = 62.09, p < .0001 \), and vowel \( F(1,61) = 20.09, p < .0001 \). A robust Group × Test interaction \( F(2,61) = 25.48, p < .0001 \) was also found, indicating a significant difference in performance between groups. Group × Vowel \( F(1,61) = 5.31, p < .0001 \) and Test × Vowel \( F(1,61) = 25.48, p = .031 \) interactions were also significant since the two trained groups improved the identification of the two vowels after training.

Post-hoc pairwise comparisons (Bonferroni) on the Group × Test interaction showed a significant difference between groups in the posttest \( F(2,61) = 26.60, p < .0001 \), but not in the pretest \( (p = .91) \). Moreover, a significant effect of test was found with both the HVPT \( F(2,61) = 94.56, p < .0001 \) and the LVPT \( F(2,61) = 17.03, p < .0001 \) groups, but not with the control group \( (p = .79) \). The interaction of Group × Vowel was significant because the HVPT group identified the vowel /æ/ better than the LVPT for 6.29% \( (p = .030) \) and the control group for 14.86% \( (p < .0001) \). The Test × Vowel interaction was also robust as the identification of both vowels from pretest to posttest improved (both with \( p < .0001 \)). The two trained groups were found to exhibit improvement of the two vowels after training (all with \( p < .0001 \)). There was no significant difference in the identification of the two vowels in the posttest by the two trained groups.

Another one-way ANOVA was computed to compare the posttest performances of all three groups. Significant differences were observed between the posttest identification accuracies of the HVPT group and the LVPT group \( (p = .002) \), the HVPT group and the control group \( (p < .0001) \) as well as the LVPT group and the control group \( (p = .004) \). It reveals that after the perceptual training, the two trained groups performed significantly better than the control group, while the HVPT group showed more significant improvement than the LVPT group. In sum, the percentage performance in the posttest phase of the two training groups was 70.36% \( (SD = 10.72\%) \) and 61.17% \( (SD = 8.96\%) \) respectively, with 19.45% of significant improvement in the HVPT group and 9.65% of improvement in the LVPT group.

Concerning the generalization effects, for TG1, a two-way ANOVA showed a significant effect of group \( F(2,61) = 24.31, p < .0001 \) and vowel /æ/ only: 14.64%, \( SD = 4.84\% \), \( p = .005 \); HVPT > Control \( /\epsilon/: 24.66\%, SD = 4.60\%, p < .0001; /\epsilon/: 27.60\%, SD = 4.52\%, p < .0001 \); LVPT > Control \( /\epsilon/ only: 16.26\%, SD = 4.72\%, p = .003 \). Both trained groups performed better than the control group, while the HVPT subjects were also more accurate than the LVPT in the TG1.

With regard to the subjects’ performance in TG2, another one-way ANOVA showed that the main effects of group \( F(2,61) = 25.685, p < .0001 \) and vowel \( F(1,61) = 7.71, p = .007 \) were significant. The LVPT group again outperformed the LVPT group by 12.503% \( \text{for } /\epsilon/ \text{ only} (p = .023) \); it also had higher mean scores than the control group by 21.29% and 24.98% \( \text{both } p < .0001 \) \text{for } /\epsilon/ \text{ and } /\epsilon/ \text{ respectively, and so did the LVPT over the control group, by 19.56% for } /\epsilon/ (p < .0001) \). Results in TG1 and TG2 showed that perceptual learning could be generalized to new words and new voices in general, although to only one vowel in some cases.

3.2. Transfer of perceptual learning to production

A three-factor repeated measures ANOVA revealed the difference between groups was significant in the posttest \( F(2,61) = 15.36, p < .001 \), but not in the pretest \( (p = .19) \). A significant effect of test was also obtained with the HVPT \( F(2,61) = 61.0, p < .0001 \) and the LVPT groups \( F(2,61) = 61.0, p = .001 \), but not with the control group \( (p = .74) \). In other words, only the HVPT and the LVPT groups’ improvements were significant: they improved from 62.05% \( (SD = 9.54\%) \) to 80.23% \( (SD = 15.86\%) \) and from 60.70% \( (SD = 9.4\%) \) to 67.63% \( (SD = 10.72\%) \).
14.67%) respectively. The significant Group × Vowel interaction was due to the consistently better production of the vowel /ε/ than /æ/ among all groups (HVPT: 29.09%, p < .0001; LVPT: 49.74%, p < .0001; control: 73.48%, p < .0001). Another one-way ANOVA showed significant difference between the posttest performance of different groups: HVPT > LVPT but only for /æ/ (23.66%, p = .034) and HVPT > control (for /æ/ only: 44.5%, p < .0001). The difference between the LVPT and the control groups for both vowels was not robust. This result may be because the subjects had reached a ceiling effect of 86.46% (SD = 13.86%) of /ε/ production in the pretest. Their performance in the posttest was only limited (posttest scores for HVPT: 93.63%, SD = 4.59%; for LVPT: 92.11%, SD = 10.44%).

Although the production accuracy in TC appeared to differ among the three groups: HVPT (mean = 72.64%, SD = 14.07%) > LVPT (mean = 67.68%, SD = 13.78%) > Control (mean = 63.30%, SD = 10.83%), the effect of group was not significant (p = .062). This suggests that the perceptual learning effect on production did not generalize to contextualized speech. Yet, the effect of vowel was still significant [F(2,61) = 186.09, p < .0001], with /ε/ produced more accurately than /æ/ (HVPT: 49.27%; LVPT: 52.00%).

4. Discussion

The results confirmed that both the HVPT and the LVPT are useful paradigms, though to different extents, in modifying the perception and production of English /æ/-/ε/ contrasts. The HVPT approach was found to be a more effective approach than the LVPT in perceptual training of the two vowels. This implies that perceptual stimuli with different phonetic environments and produced by multiple talkers are more useful in shaping the perceptual acquisition of the subjects, as stimulus variability promotes learners’ selective attention to the acoustic cues which are to be observed by the subjects from the large pool of stimuli with wider variability. Exposure to highly variable stimuli is necessary for the subjects to form robust phonetic representations by learning which acoustic cues are relevant to a specific sound; that is, when the subjects were allowed to selectively attend to a wider range of acoustic dimensions and weightings, they developed more language-specific phonetic categories. This finding is in line with previous HVPT studies. The fact that perceptual learning was successfully transferred to production, with the HVPT group outperforming the LVPT group also hints at a common mental representation underlying both perception and production, although at this preliminary stage no definite conclusion can be drawn about whether it supports any hypothesized mechanisms.

Previous studies showed that the degree of generalization to the perception of new words produced by both familiar and unfamiliar talkers to subjects trained under the LVPT was far less than those under the HVPT, and this has been confirmed in the present study. The findings in TG1 and TG2 suggested the influential role of stimulus variability in the generalization process, which requires more language-general mapping of linguistic categories. The greater success of generalization of the HVPT group is attributable to the subjects’ focus on the criterial properties and acoustic cues in common of the vowels produced by the different speakers, in which they also had to strive to ignore the between-speaker variability that might present obstacles to the perceptual learning. The HVPT subjects can also draw from a larger pool of tokens stored in their memory during identification. In contrast, the LVPT subjects were only exposed to a single speaker and hence could have only knowledge of one speaker to apply to their perceptual strategies in identifying new words from new speakers. Their encoding strategies were shallower and more limited than that of the HVPT group. Research investigating memory and categorization also suggests that training under high variability conditions could contribute to more robust generalization [8]. Yet, the findings for the production contextualization data indicate that even though both training groups outperformed the control group, the effect was not significant. This may suggest that brief phonetic training at the word-level was not sufficient to promote production improvement at the sentence level, and follow-up studies are necessary to test whether a lengthened, more intensive, or sentence-level training will be more beneficial to learners.

5. Conclusion

The findings have provided preliminary empirical evidence that training in perception alone can be useful for improving both the perception and production of the non-native contrast among Cantonese speakers of English, with stimulus variability playing a crucial role in the amount of improvement. Positive generalization effects to the perceptual domain suggest that the efficacy of the HVPT is externally valid. This study, by comparing and contrasting directly the HVPT and the LVPT paradigms, also confirms previous studies and offers more solid support to the claim that stimulus variability is at work under laboratory-based perceptual training. However, investigating only one non-native contrast and lacking native speakers’ judgment on subjects’ production data are the limitations of the study. Future studies should draw on a wider range of phonetic contrasts and have native speakers’ evaluation. Other variations such as language proficiency as a factor, sentence-level phonetic training or even whether production training can also affect perception, would also be relevant foci for future studies.

6. References