The Impact of L2 Proficiency on Vowel Training

Wong, Janice W S

Published in:
Investigating English Pronunciation

DOI:
10.1057/9781137509437_10

Published: 01/01/2015

Document Version:
Peer reviewed version

Wong, J. W. S. (2015). The Impact of L2 Proficiency in Vowel Training. In J. A. Mompean & J. Fouz-González (eds.), *Investigating English pronunciation: Current Trends and Directions* (pp.219-239). Hampshire: Palgrave Macmillan.

The Impact of L2 Proficiency on Vowel Training

Janice Wing Sze Wong
Hong Kong Baptist University

Abstract

The present research aimed to investigate the effect of English proficiency level on Cantonese ESL learners' learning of an English vowel contrast (namely, /e/ and /æ/) through a perception-based training paradigm called High Variability Phonetic Training (HVPT) which uses perceptual stimuli in multiple phonetic environments produced by various speakers (e.g. Logan, Lively & Pisoni, 1991). All subjects participated in both perception and production pretests and posttests. Twenty-two (9 high proficiency; 13 low proficiency) subjects were trained under 10 sessions of HVPT in which they had to identify /e/ and /æ/. Another 23 subjects (10 high proficiency; 13 low proficiency) served as control group. The results showed that exposing low and high proficiency learners to highly-variable natural stimuli can successfully train the perception of a non-native phonetic contrast. Transfer of perceptual learning to production and generalization of learning were also observed. The results obtained suggest that L2 teachers should consider adopting similar training paradigms in classrooms.

Keywords: Speech perception and production, vowel training, second language acquisition, High Variability Phonetic Training

1. Introduction

Previous studies have shown that the perception and production of non-native contrasts are difficult even for advanced second language (L2) learners (Bohn 1998; Polka 1992; Strange 1995). Various types of L2 phonetic training paradigms have been devised either to investigate the relationship between speech perception and production and thus benefit the theoretical field, or to benefit both teachers and learners practically (e.g. perception-only training: Bradlow et al. 1997; Lambacher et al. 2005; production-only training: Carruthers 2007; Hattori and Iverson 2008; Leather 1997; perception training involving audio-visual materials: Aliaga-García and Mora 2009; Hazan et al. 2005; perception-and-production training: Tsushima and Hamada 2005).

Among all types of phonetic training paradigms, High Variability Phonetic Training (HVPT) which was firstly proposed by Logan and colleagues (1991) has received particular attention in recent decades due to its effectiveness in improving the perception and production of different consonant and vowel contrasts by L2. HVPT involves the use of natural training stimuli with various phonetic contexts produced by multiple speakers. It is usually administered in the form of a discrimination or identification task with immediate feedback.

HVPT has been used as a basic framework of phonetic training and its different variations have proved to be successful (e.g. Iverson and Evans 2007; Nishi and Kewley-Port 2005, 2007; Sperbeck et al. 2005). Significant improvement can be found among many subjects with different L1s after training (e.g. Flege 1995; Hirata 2004; Hirata et al. 2007; Kim and Hazan 2010; Pruitt 1995; Pruitt et al. 2006; Rochet 1995; Tajima et al. 2008; Wang and Munro 1999; Yamada et al. 1996). In addition, using highly variable training stimuli has been found to promote the perceptual learning of the subjects (e.g. Bradlow et al. 1997; Bradlow et al. 1999; Lively et al. 1993; Lively et al. 1994; Logan et al. 1991), and generalization effects to new words and new speakers have also been also obtained, especially when the subjects were trained with a wider range of stimulus variability (e.g. Bradlow et al. 1997; Wang 2002; Wong 2013; 2014b). Several extended studies (e.g. Bradlow et al. 1997; Bradlow et al. 1999; Lively et al. 1994) have also shown that the training effects can be retained in the long run. Perceptual learning through HVPT has also proved to be capable of transferring to the production domain, although a wide range of individual differences among learners have been observed across studies (e.g. Bradlow et al. 1999; Hazan et al. 2005; Lambacher et al. 2005).

2. Phonetic training and proficiency

Despite the various benefits that HVPT seems to provide, most of these previous training studies have only tested advanced adult L2 learners who still had problems in perceiving and/or producing some segmental contrasts, overlooking the training effects on low proficiency learners. The studies seldom consider proficiency as a possible extraneous factor that may influence the degree of learning of the subjects and hence the external validity of the training paradigm. In fact, language proficiency has remained as a largely unexplored area compared to well-documented factors such as age of L2 learning (e.g. Akahane-Yamada 1995; Flege et al.

1999) or phonological inventories of the L1 and L2 (e.g. Kuhl 2000; Polka 1991). Studies have usually investigated the factor L2 proficiency indirectly. They examined related notions such as phonological short-term memory (e.g. Hummel 2009; MacKay et al. 2001), pitch-level ability (Lee et al. 2007, 1589-1592) or even musical ability (e.g. Alexander et al. 2005, 397-400; Slevc and Miyake 2006) on L2 perception. These factors have been shown to influence the success in L2 perception. The present study hopes to shed light on this aspect by involving a more general variable: the general L2 proficiency in perception (evaluated by their grades in a listening public exam) and production (evaluated by their grades in an oral public exam). The aim is to shed light on how language proficiency may influence L2 speech perception and production in general.

The present research focuses on the perceptual learning and transfer of learning to production of the English /e/-/æ/ vowel contrast among Hong Kong Cantonese ESL learners[1]. This vowel pair was chosen because a number of studies on the perception and production performance of English vowels have indicated that this vowel pair poses production and perception problems for Hong Kong Cantonese speakers (e.g. Chan 2010, 2012; Chan and Li 2000; Hung 2000; Leung and Brice 2012; Meng et al. 2007; Sewell 2009). Hung (2000), for example, ascribed the problem to the differences between the L1 (Hong Kong Cantonese) and L2 (English) phonological systems. The L2 realizations of /e/ and /æ/ have also been found to cause intelligibility problems for native speakers (Brown 1991; Jenkins 2000; Sewell 2009). This chapter presents the results of an experiment on the extent to which proficiency levels affect perception and production, and whether or not L2 vowel training paradigms can benefit L2 learners. The research questions are as follows:

RQ1. Is HVPT effective in improving native Cantonese ESL learners' perception and production of the English vowels /e/ and /æ/?

---

[1] Portions of the present research (paper written in 2013) have been published in two conference proceedings focusing on different aspects (Wong, 2012; Wong, 2014a). Wong (2012) compares the results of the same group of subjects receiving HVPT with a group of subjects receiving another training paradigm, LVPT. Wong (2014a) is a later report on the interaction of the effects of stimulus variability and learners' proficiency level. The present article focuses on the link between perception and production by comparing the degree of learning in the two domains. This paper also provides a deeper investigation in the production data of the subjects by offering acoustic analyses on the production data (not just transcription results) before and after training.

RQ2. If HVPT is effective, can the effects of training be generalised to the perception of new words with /e/ and /æ/ produced by both familiar and new speakers, or to the production of the two vowels in a more naturalistic environment?

RQ3. What are the effects of English proficiency in training the subjects' perception and production of the English vowels /e/ and /æ/?

2.1. Methodology

Participants

A total of 45 Hong Kong secondary school students with Cantonese as their first language participated in the experiment. Their average age was 17.1 (*SD* = 0.63) and average age of learning English as an L2[2] was 3.1 (for an average of about 14 years, *SD* = 0.52). They all shared similar English-learning background and amount of exposure to English, which was reflected in a survey they were asked to complete at the beginning of the study. None had resided in any English-speaking countries. None reported hearing or speaking problems.

Twenty-two of them (12 females and 10 males) were trained under HVPT. The experimental group was further divided into two groups: 9 with a high English proficiency level (5 females, 4 males) and 13 with a low proficiency level (7 females, 6 males). A control group with 23 participants (12 females and 11 males) also participated in the pre/posttests without training. Ten of them had a high proficiency level (5

---

[2] English is regarded as one of the official languages in the Hong Kong SAR Government and is widely used in commerce, administration and education. English is learnt as an L2 standard by Hong Kong locals and formal learning of English begins as early as the age of three. The concept of *English as an L2 in Hong Kong* is not to be confused with the term *Hong Kong English*. The existence of Hong Kong English still remains controversial particularly as far as its grammar and usage are concerned. The administration and business sectors do not accept this alleged "variety of English" and only deem its users as incompetent English learners. However, it is evident that most Cantonese speakers of English produce a vowel set different from that of native speakers of English (as those in the Inner Circle, Kachru 1985) in general and they speak with an accent which is characterized by a flat intonation, a syllable-timed rhythm and simplified consonant clusters, etc. Local Hong Kong people generally call this a "Hong Kong accent."

females, 5 males) whereas 13 had a low proficiency level (7 females, 6 males). Table 1 illustrates the grouping dichotomy:

Proficiency level / Group	High	Low	Total
HVPT	9	13	22
Control	10	13	23

Table 1. Number of participants in each group.

The participants' English proficiency levels were determined by the average grades obtained in the listening and oral papers in a recognized Hong Kong public exam named *Hong Kong Certificate Education Examination* (HKCEE), a reliable and reflective assessment of their English proficiency. Participants from the high proficiency level group were from the top 5.8% of the entire candidate pool in 2009 - a total of 109,135 candidates in that year- in Hong Kong; those with low proficiency level were from the lower 60% (HKEAA, 2009). This experiment took place 4 months after the subjects finished the 2009 HKCEE.

Stimulus Materials used in Perceptual Training

Six native General American English speakers, aged 30-45, recorded all the stimuli in perceptual pre/posttest and training. In total, they contributed 60 /e/ and /æ/ real word minimal pairs (6 speakers, each produced only 10 tokens) under various consonant-vowel-consonant (CVC) environments. These are the stimuli used in the training. One of the six native speakers also recorded all the test tokens and 10 distracters used in the perception pre/posttests. Another one among the six, i.e. a familiar speaker to the subjects, recorded another word list with 30 /e/ and /æ/ minimal word pairs for TG2 (new words by a familiar speaker). Recordings from an additional speaker who had not recorded anything previously, i.e. a new speaker to the subjects, were also obtained. This seventh recorded another 30 /e/ and /æ/ new minimal pairs for TG1 (new words by a new speaker).

In order to avoid intra-speaker variability in vowel productions (Wang, 2002), each speaker read the tokens at least three times so that no single token would be used for all stimuli. All three tokens were evenly and randomly used in the training program to avoid speaker effect reported in earlier studies (e.g. Logan et al. 1991; Lively et al. 1993; Lively et al. 1994).

Procedure

The study involved three phases: pretest, perceptual training and posttest. The experimental group completed both the pre/posttests and training whereas the control group did only the tests.

During the pretest phase, all subjects completed one production pretest and one perception pretest. In the production pretest, the subjects had to record a wordlist of 60 items (30 /e/ and 30 /æ/) and 10 distractors. No audio prompts were provided during the recording. In the perception pretest, the subjects used the computer program to complete an identification test. 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. The frequency of occurrence of the correct answer in the four serial positions was equal, allowing the chance level to be fairly inferred at 25%. The subjects could play the audio clips multiple times according to their own needs before submitting their answer. Figure 1 shows two screenshots of the computer program.

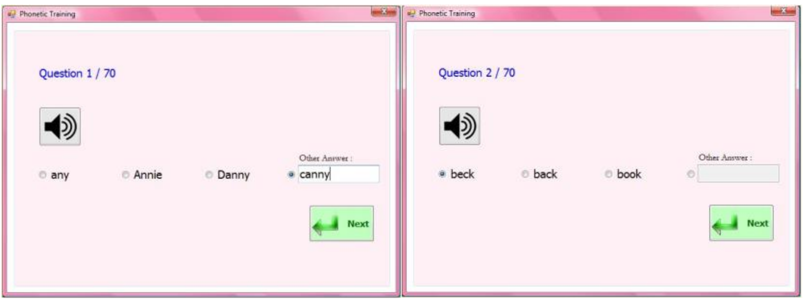

Fig. 1. Screenshots of the perceptual pretest

During the perceptual training phase, the experimental group participated in a total of 10 HVPT training sessions over five days (i.e. two sessions per day). The participants were presented with 70 stimuli produced by six native English speakers and were trained on a two-alternative forced choice (2AFC) paradigm to ensure training intensity, as subjects only had to focus on one word at a time and choose between two options. This was different from offering them four choices in the tests, although the 4AFC test design was aimed to achieve more accurate reflection of the subjects' performance by lowering the percentage of identification based only sole speculation from 50% to 25%. Again, subjects could repeatedly play the audio clips before confirming their answer. During training, immediate feedback was given and they could listen to the sound clip again before they confirmed their choice; at the end of each session, their total scores were also shown. Figure 2 illustrates a series of screenshots of the training program:

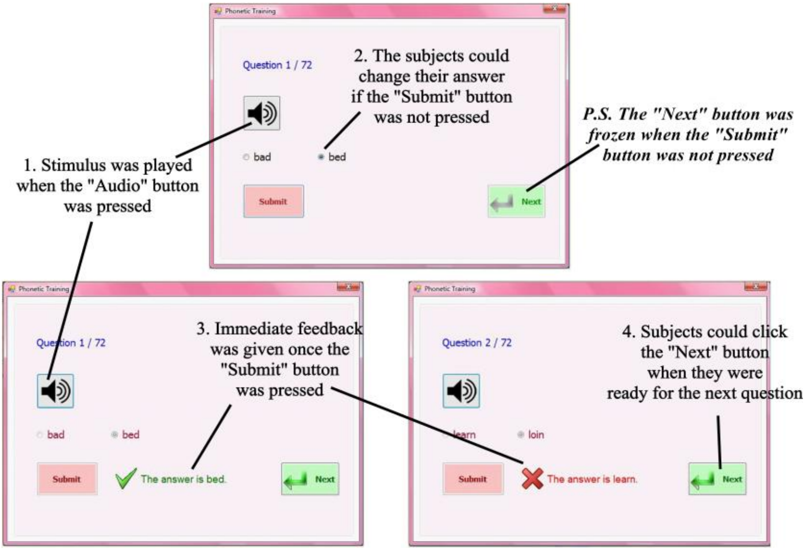

Fig. 2. Screenshots of the training program

Finally, subjects completed a post-test. In this phase, all subjects completed one production posttest, one perception posttest, one production Test of Contextualization (TC) and two perception Tests of Generalization (TG1 and TG2). Both production and perception posttests were the same as those in the pretest phase. For the TC, all subjects recorded a 250-word passage including 50 items with the target vowel pairs. Perceptual TG1 and TG2 involved the use of 30 new words spoken by a new speaker and 30 new words spoken by a familiar speaker respectively. The procedures were the same as those in perception pre/posttests.

All sessions took place in a quiet language laboratory. Perceptual tests and training were completed using a computer program designed by the researcher, with stimuli played over headphones at a comfortable volume. The subjects could adjust the volume based on their own needs. All the subjects were assigned a participant ID and password with which they could log in to a specific account of the computer program. Through this, the researcher could also ensure that all sessions were completed by a particular subject. The production test tokens were recorded using *Adobe Audition 1.5* in separate partitions.

Evaluation of Data

All the data of the production tests were transcribed twice by the researcher, a native speaker of Cantonese with English as L2. The researcher adopted a two-way (correct-incorrect) scoring system during the transcription. The intra-rater reliability was calculated by using the total number of target productions produced by all the subjects in the second transcription divided by the first trial of transcription. The reliability was 94.8% with Cronbach's alpha $\alpha = .832$. A second, phonetically-trained, native English-speaking researcher was also invited to check the inter-rater reliability by doing 50% of the transcriptions. The inter-rater reliability reached 91.6% with Cronbach's alpha $\alpha = .802$.

The data were also analysed acoustically by using the *Praat* speech analysis software (Boersma and Weenink 2005). The first two formant frequencies (F1 and F2) and the temporal measurements of the vowels were gauged at the midpoint to evaluate how similar or different the vowel productions were after training.

2.2. Results

Perceptual performance (pretest vs. posttest)

The results of the identification performance of the HVPT and control groups with a high and low proficiency levels are displayed in Figure 3. A three-factor ANOVA was computed using Group (HVPT, control), Proficiency (high, low) and Type of Test (pretest, posttest) as factors. It showed a significant main effect of Group [$F(1,43) = 24.05$, $p < .001$] and Type of Test [$F(1,43) = 50.95$, $p < .001$]. The HVPT group (high proficiency) showed an improvement of 21.85% ($p < .001$) whereas the low proficiency group showed a gain of 17.05% ($p < .001$) from pretest to posttest. A robust Group × Type of Test interaction [$F(2,43) = 55.213$, $p < .001$] was also found, indicating a significant difference in performance between groups. Post-hoc pairwise comparisons (Bonferroni) on the Group × Type of Test interaction showed a significant difference between groups in the posttest [$F(2,43) = 56.34$, $p < .001$], but not in the pretest ($p = .75$). Moreover, a significant effect of Type of Test was found within the HVPT groups [$F(1,43) = 102.98$, $p < .001$], but not with the control groups ($p = .83$). However, Proficiency was not a significant factor ($p = .82$) and neither were the interactions Type of Test × Proficiency ($p = .23$), Group × Proficiency ($p = .29$) and Type of Test × Group × Proficiency ($p = .58$).

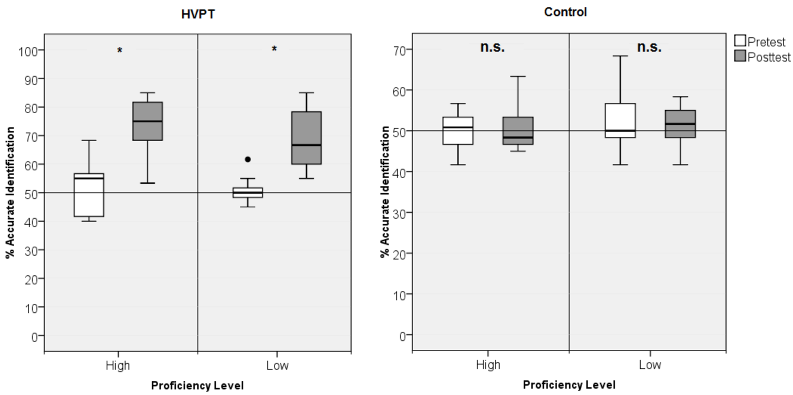

Fig. 3. Boxplots of identification accuracy of the two target vowels in the pretest (white boxes) and the posttest (dark boxes); HVPT (left) and control (right) groups were subdivided into two proficiency levels. The horizontal line indicates the chance level performance. The asterisks represent that the difference is significant whereas n.s. represents non-significant difference.

Generalization of learning in the perceptual domain

Figure 4 displays the identification accuracy of both the experimental and control groups with different proficiency levels in two different generalization tests, TG1 and TG2. A two-way ANOVA with Group (HVPT, control) and Proficiency (high, low) showed only a significant main effect of Group in both TG1 [$F(1,43) = 6.375$, $p < .001$] and TG2 [$F(1,43) = 3.46$, $p = .002$], since the experimental groups performed significantly better than the control groups. Yet, Proficiency was not a robust factor (TG1: $p = .53$; TG2: $p = .74$).

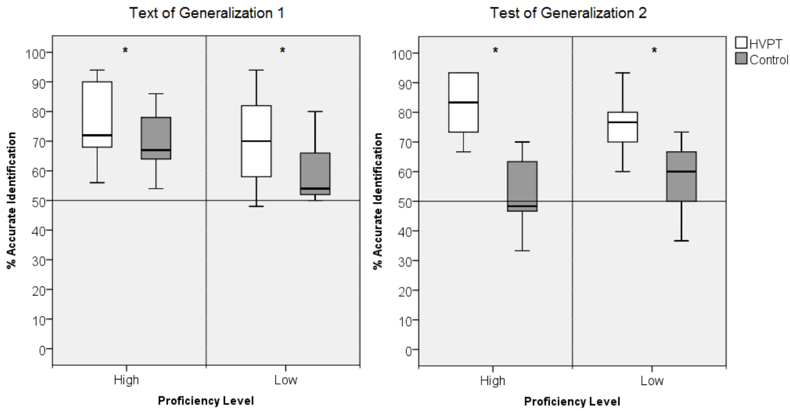

Fig. 4. Boxplots of identification accuracy for the two target vowels in TG1 (left) and TG2 (right). The white boxes are the performance of the HVPT group whereas the dark boxes represent the performance of the control group. The horizontal line indicates the chance level performance. The asterisks represent that the difference is significant.

Transfer of learning to the Production Domain (Pretest vs. Posttest) – based on transcription results

The results of the production performance of the groups in the pretest and posttest are shown in Figure 5. A three-factor repeated measures ANOVA with Group (HVPT, control), Proficiency (high, low) and Type of Test (pretest, posttest) as factors showed significant main effects of Group [$F(1,43) = 17.06$, $p < .001$], Type of Test [$F(1,43) = 79.51$, $p < .001$] as well as a significant interaction of Group × Type of Test [$F(2,43) = 70.17$, $p < .001$], since the HVPT groups improved their production performance from pretest to posttest for 21.48% (high proficiency group) and 15.90% (low proficiency group). Post-hoc pairwise comparisons (Bonferroni) on Group × Type of Test interaction showed that the effect of Test was significant between groups in the posttest [$F(1,43) = 32.41$, $p < .001$] but not in the pretest ($p = .11$). In addition, the effect of Group was also significant in the HVPT groups [$F(1,43) = 145.21$, $p < .001$], but not in the control groups ($p = .70$). Again, Proficiency was not a significant factor ($p = .15$), and neither were the interactions Type of Test × Proficiency ($p = .19$), Group × Proficiency ($p = .65$) and Type of Test × Group × Proficiency ($p = .21$).

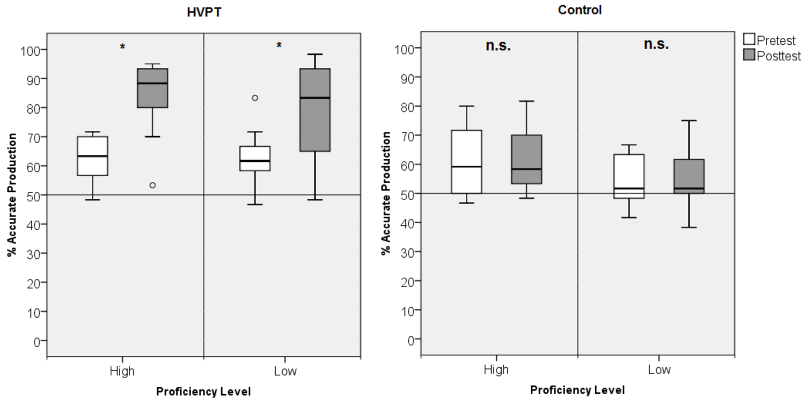

Fig. 5. Boxplots of production accuracy for the two target vowels in the pretest (white boxes) and the posttest (dark boxes), for the experimental (left) and control (right) groups which are subdivided into two proficiency levels. The horizontal line indicates the chance level performance. The circles indicate the outliners. The asterisks represent significant differences whereas n.s. represents non-significant difference.

Production performance in passage reading task (TC) – based on transcription results

Figure 6 displays the production accuracy in the passage reading task, Test of Contextualization (TC). A two-way ANOVA with Group (HVPT, control) and Proficiency (high, low) showed that both the effects of Group (p = .37) and Proficiency (p = .50) were not significant, suggesting that the perceptual learning effect on production did not generalize to contextualised speech.

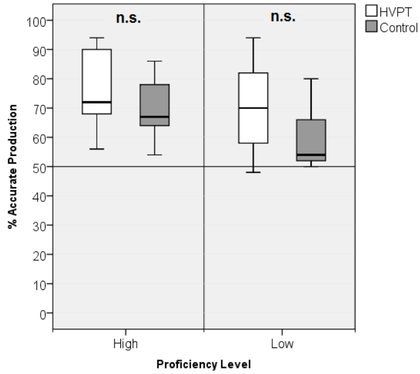

Fig. 6. A boxplot of production accuracy for the two vowel pair in TC, a passage reading task. The white box shows the performance of the HVPT group whereas the dark box represents that of the control. The horizontal line indicates the chance level performance. The notation n.s. represents non-significant difference.

Acoustic analysis on production pre/posttest data

As mentioned above, the first two formant frequencies (F1 and F2) and vowel duration of the words produced by both the native speakers and the subjects in the pretest and posttest were measured with *Praat*. The formant frequency measurements of each vowel measured at the vowel midpoint were estimated by the formant tracking function in *Praat*. From dual spectrogram and waveform displays, the temporal measurements of the vowels were also measured. The mean F1 and F2 values as well as the vowel duration of /e/ and /æ/ produced by both the native speakers and the subjects are shown in Figures 7 and 8 respectively.

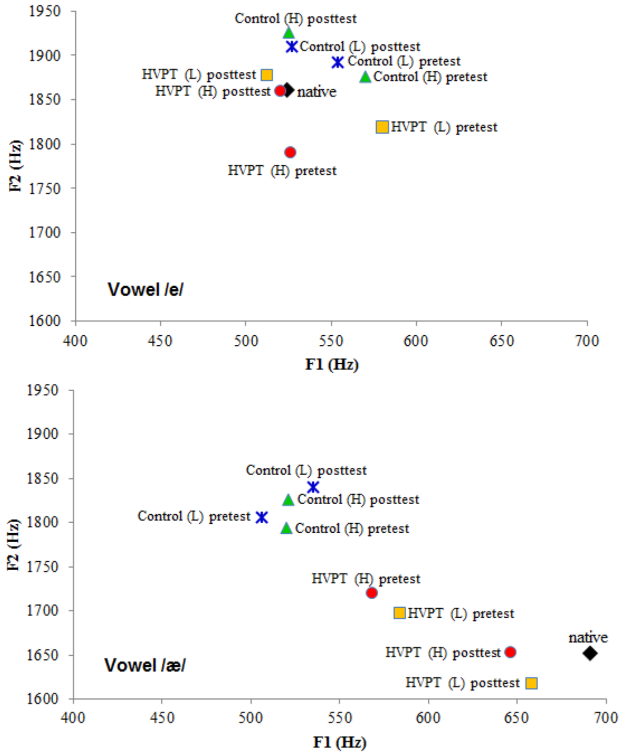

Fig. 7. The mean values of F1 and F2 of the two target English vowels produced by different groups in production pretest and posttest.

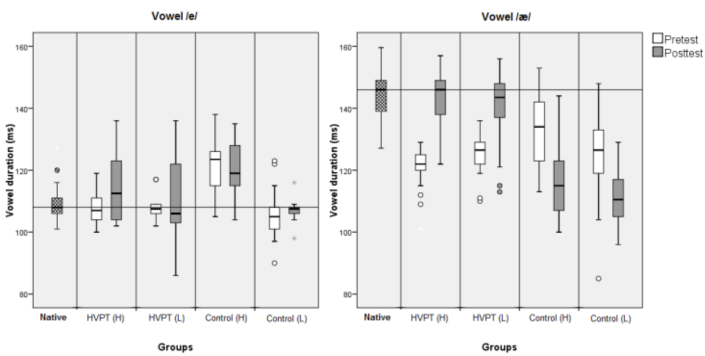

Fig. 8. Durations of vowels produced by the native speakers and the four groups of subjects (the horizontal line represents the mean values of the native speakers' productions).

For the first formant of vowels, all groups in the pretest produced both vowels with F1 values closer to the vowel /e/. After training, the F1 values for words with the vowel /e/ still remained similar to those produced by the native speakers, indicating that the production of this vowel in terms of F1 were fairly close to native performance. The HVPT groups also produced the vowel /æ/ with higher F1 which are closer to those produced by native speakers. However, the F1 value of the vowel /æ/ produced by the control group in the posttest remained close to those in the pretest. This suggests that a considerable amount of subjects in the training groups produced the vowel /æ/ with a more open vowel height, which is one of the articulatory differences of the production between the vowels /e/ and /æ/.

Considering the second formant frequency, all the groups produced the two vowels in the pretest with values around the same frequencies, which are also similar to the F2 values of the vowel /e/ produced by the native speakers. However, in the posttest, both training groups produced more native-like F2 frequencies for both vowels, meaning that the subjects started to be able to distinguish the production of the two front vowels after training.

Regarding vowel duration, both experimental groups produced the two vowels with very similar vowel duration in the pretest. The duration ratio of /e/ to /æ/ in the pretest for the HVPT groups were 1.10 (high proficiency) and 1.12 (low proficiency) and the control group was 1.06 (high proficiency) and 1.21 (low proficiency), all noticeably lower than that of the native speaker group (1.34). After training, the ratio for the control groups was still similar, but the HVPT groups produced the two vowels with larger duration ratio (high proficiency = 1.27; low proficiency = 1.29), suggesting that their production had become more native-like.

Linking up perception and production data

Besides looking at the results in the perception and production tests separately, comparing and contrasting the amount of learning in the perception and production pre/posttests of different groups can also shed light on the relationship between the two domains and the degrees of learning of the subjects. This comparison can be made using a representation of perceptual-production space of each participant (Bradlow et al. 1997). All vector plots of the groups, categorized under their proficiency levels are presented in Figure 9.

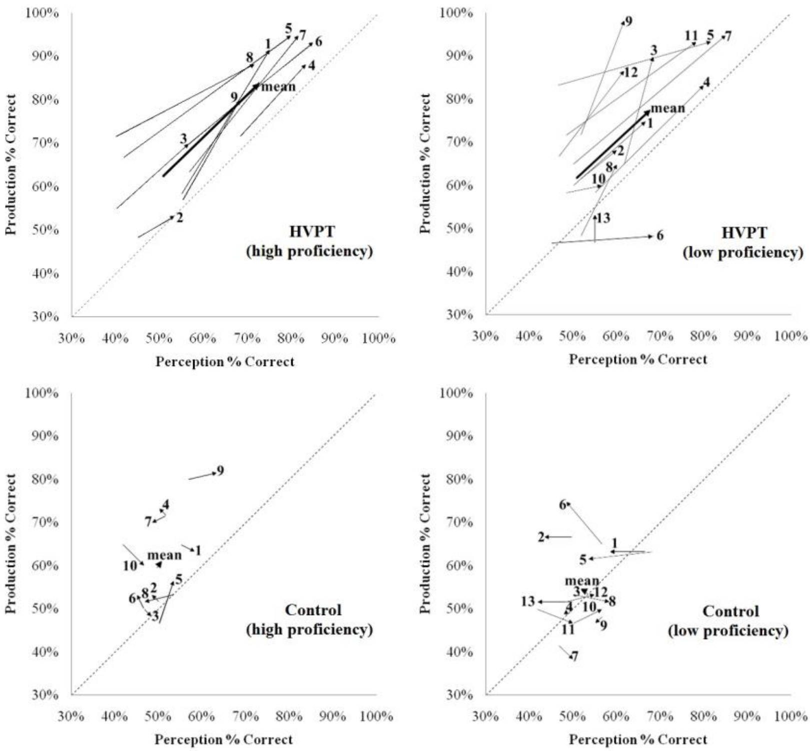

Fig. 9. Vector plots of individual subjects' perceptual identification accuracies (*x* axis) and target productions (*y* axis) from the pretest to the posttest. A numbered vector is used to indicate each individual's performance. The bold arrow represents the group mean, while the dotted diagonal indicates the hypothetical and ideal location for a perfect correlation between speech perception and production.

The perceptual-production space of each group illustrates the amount of learning of all individual subjects in both domains. Their performance is indicated by the vector, with the direction of arrow signalling the change from pretest to posttest in both perception and production dimensions. The bold arrow shows the mean percentage of the group while the diagonal shows the ideal direct proportional change which correlates the change in perception and production.

Previous studies (e.g. Bradlow et al. 1997; Hazan et al. 2005; Wang 2002) have shown that there are frequent individual differences in cross-sectional studies examining the perception and production of non-native contrasts. Figure 7 above evinced that the two groups with the HVPT training showed fairly noticeable improvement in both domains since most of the subjects showed a long vector indicating a change of over 20% after the training. Yet, the performance of the control group subjects does not follow a regular pattern. Most of them had only minor changes (shown by the relatively short vectors) in their performance.

The general and consistent success of the subjects in the HVPT groups already suggests that perceptual learning also leads to improvement in the production domain, provided that the subjects are given sufficient and highly variable stimuli in the identification tasks. However, learning in the two domains proceeded at different rates. Generally, similar positive slopes and relatively great length of the vectors also provide evidence at a possible link between speech perception and production may exist since a considerable amount of perceptual learning did transfer to production. The correlation of the degree of learning between the two domains was fairly high and a preliminary conclusion could be reached: perceptual training under the HVPT approach leads to a larger amount of improvement in the production of the subjects.

2.3 Discussion

The results showed that, in general, exposure to highly variable stimuli was helpful for the subjects to improve their perception of a non-native contrast. It may probably be due to the formation of robust phonetic representations as they could learn which acoustic cues were relevant to a specific sound. When they were allowed to selectively attend to a wide range of acoustic dimensions and weightings, they developed more language-specific phonetic categories. Perceptual generalisations to new words/speakers also took place because the subjects could also focus on the criterial properties and acoustic cues that were common in vowels produced by different speakers; hence, they gained good encoding strategies. Moreover, learning in the perceptual domain could also be transferred to production, suggesting the possibility that there may be a common mental representation underlying both perception and production.

Previous studies (e.g. Bradlow et al. 1997; Strange and Dittmann 1984; Wang 2002; Akahane-Yamada 1995, 305-320) investigating the effectiveness of HVPT in the modification of both/either the perceptual and/or the production of English consonants or vowels showed substantial individual variations in the degree of acquisition in the two domains. This can also be found in the variations of the vectors shown above. The analyses further implied that there were some underlying factors affecting individual performance. The results showed that both high and low proficiency groups benefited from the training and with performances that were not significantly different from each other. These findings are different from previous studies related to the investigation of language proficiency in general, which showed a positive correlation or influence to L2 learning: the more proficient the subjects are, the better results they will gain. The perception and production of L2 sounds are still a common difficulty among L2 learners even if they have acquired a high level of proficiency in other areas such as L2 reading/writing.

With a view to providing a fair ground for comparison, the present study employed and limited the notion of L2 proficiency to the subjects' general proficiency in their language perception and production levels through the evaluation of their listening and oral exam results. The present data suggest that proficiency level did not influence the performance of the subjects. It was speculated that L2 phoneme perception and production may require a higher-level of processing that involves the utilisation of an innate and human-specific specialised module but not simply general listening or speaking ability. A higher ability in general listening of the L2 may not lead to the acquisition of minor acoustical or durational cues of the vowels. Also, high proficiency level in general speaking and communication in the L2 may not be necessarily linked to finer articulatory and motor gestures. Moreover, the listening and oral exam in Hong Kong which was used in the present study for judging the proficiency level may require more intelligence, communicative ability and general understanding of the contexts but not simply speech perception and production ability of particular contrasts. These may constitute some other confounding variables, leading to the present finding.

3. Conclusion

The main aim of the present research was to determine whether the level of English proficiency would be a determining factor in improving the perception and production of the English /e/-/æ/ contrast by Cantonese ESL learners under a perceptual-based phonetic training paradigm called High Variability Phonetic Training. HVPT is characterized by the adoption of perceptual stimuli produced by multiple speakers and with multiple phonetic environments. Under a pretest-treatment-posttest design, the present experiment recruited a total of 45 Cantonese ESL learners**.** Twenty-two of them completed 10 sessions of HVPT in which they identified the vowel they heard and the remaining subjects served as control group. The results confirmed that HVPT is a useful paradigm that enhances the perception and production of English /e/-/æ/ contrasts by Cantonese ESL speakers with different English proficiency levels.

Given the promising results in this study, it appears that the difference in general English (listening and speaking) proficiency levels did not affect learning in specific auditory training. The results also provided preliminary evidence that HVPT was effective for improving the perception and/or production of non-native contrasts of learners in general. As HVPT is a paradigm that can be implemented and set up easily inside or outside the classroom, language teachers could consider adopting this auditory training into their instructional proposal to allow the learners to have more opportunities to be exposed to contrasts that may result in serious communication breakdown due to the lack of intelligibility caused by the conflation of a contrast. This can be done with software specifically designed for this purpose (as in the case of the present research), although this can be extremely challenging for the teachers. Thus, free software like the one created by Rato and colleagues (Rato, Rauber, Kluge & Santos, this volumen) holds great promise for implementing this type of training in the classroom, as it allows teachers to create their own tasks without specific knowledge of programming.

The current study has limitations that have to be acknowledged. This study adopted only identification tasks in testing and training which were different from category discrimination tasks (e.g. Iverson et al. 2012). We could only speculate that it may be the nature of the identification tasks that allows more focused attention on the phonetic difference in the target contrast, rather than the highly variable nature stimuli alone. It would be important to compare the differences between identification and category discrimination tasks in order to understand further what levels of phonological encoding or which aspects of L2 vowel perception the tasks are tapping into.

Additionally, this research did not gauge the long-term retention effect of HVPT. This is yet an important variable that should be taken into account in future studies. Given that the ultimate goal is to benefit language teachers and learners, a training program will be useless if only immediate or short-term effects are obtained.

Moreover, the acoustic analyses in the present study did not separate the results of the two genders and whether the consonant following the vowel was voiced or voiceless; rather, only the average across the two categories was reported. Future analyses must take these variables into account.

Future studies should investigate the effectiveness of HVPT into other segmental (e.g. consonant clusters or other vowel contrasts) and suprasegmental (e.g. intonation, stress or accent) elements posing difficulties to the subjects or having different phonetic realizations in Cantonese and English. Comparing the efficacy of different training paradigms can also benefit teachers and learners. Of further empirical and pedagogical interest would be to extend the investigation of the training effects to other populations of different ages, L2 experience and gender. Future research can also be oriented to the investigation of the link between perception and production by viewing what kind of training approaches can bring higher rates of improvements in both domains. All of these research extensions can further test the external validity of the training approach and offer pedagogical and theoretical values to the research field.

References

Akahane-Yamada, Reiko. 1995. "Age and acquisition of speech language sounds: Perception of American English /r/ and /l/ by native speakers of Japanese." In *Speech perception and linguistic experience: Issues in cross-language research*, ed. Winifred Strange, 305-320. Baltimore, MD: York Press.

Alexander, Jennifer A., Patrick C. M. Wong, and Ann R. Bradlow. 2005. "Lexical tone perception in musicians and nonmusicians." In *Proceedings of Interspeech 2005 - 9th European Conference on Speech Communication and Technology*, ed. Isabel Trancoso, 397-400. Lisbon, Portugal: ISCA.

Aliaga-García, Cristina and Joan C. Mora. 2009. "Assessing the effects of phonetic training on L2 sound perception and production." In *Recent Research in Second Language Phonetics/Phonology: Perception and Production*, ed. Michael A. Watkins, Andreia S. Rauber and Barbara O. Baptista, 2-31. Newcastle upon Tyne: Cambridge Scholars Publishing.

Boersma, Paul, and David Weenink. 2005. *PRAAT: Doing phonetics by computer*. Retrieved from http://www.fon.hum.uva.nl/praat/.

Bohn, Ocke-Schwen. 1998. "Linguistic relativity in speech perception: An overview of the influence of language experience on the perception of speech sounds from infancy to adulthood." In *Evidence for Linguistic*

Relativity, ed. Susanne Niemeier and René Dirven, 1-28. Amsterdam and Philadelpha: John Benjamins.
Bradlow, Ann R., Reiko Akahane-Yamada, David B. Pisoni, and Yoh'ichi Tohkura. 1999. "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production." *Perception and Psychophysics* 61(5): 977-985.
Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production." *Journal of the Acoustical Society of America* 101(4): 2299-2310.
Brown, Adam. 1991. *Pronunciation Models.* Singapore: University Press.
Carruthers, Steven W. 2007. "Articulatory Training on Facial Movements Using the Webcam Pronunciation Mirror: A Pilot Study." In *HPU TESL Working Paper Series* 5(1): 3-17.
Chan, Alice Y. W. 2010. "Advanced Cantonese ESL learners' production of English speech sounds: Problems and strategies." *System* 38(2): 316-328.
Chan, Alice Y. W. 2012. "The Perception of English Speech Sounds by Cantonese ESL Learners in Hong Kong." *TESOL Quarterly* 45(4): 718-748.
Chan, Alice Y. W., and David C. S. Li. 2000. "English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems." *Language Culture and Curriculum* 13(1): 67-85.
Flege, James E. 1995. "Second-language speech learning: Theory, findings and problems." In *Speech perception and linguistics experience: Issues in cross-language research*, ed. Winifred Strange, 233-272. Timonium, MD: York Press.
Flege, James E., Ian R. A. Mackay, and Diane Meador. 1999. "Native Italian speakers' perception and production of English vowels." *Journal of the Acoustical Society of America* 106(5): 2973-2987.
Hattori, Kota and Paul Iverson. 2008. "English /r/-/l/ pronunciation training for Japanese speakers." *Journal of the Acoustical Society of America* 123(5): 3327.
Hazan, Valerie, Anke Sennema, Midori Iba, and Andrew Faulkner. 2005. "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English." *Speech Communication* 47(3): 360-378.
Hirata, Yukari. 2004. "Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts." *Journal of the Acoustical Society of America* 116(4 Pt 1): 2384-2394.

Hirata, Yukari, Elizabeth Whitehurst, and Emily Cullings. 2007. "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates." *Journal of the Acoustical Society of America* 121(6): 3837-3845.

Hong Kong Examinations and Assessment Authority (HKEAA). 2009. "PRESS RELEASE 2009 Hong Kong Certificate of Education Examination Results Released." Accessed January 2013. Last modified August 2009. Available at http://www.hkeaa.edu.hk/DocLibrary/Media/MEM/2009_HKCEE_PressRelease_Full_ENG.pdf.

Hummel, Kirsten M. 2009. "Aptitude, phonological memory, and second language proficiency in nonnovice adult learners." *Applied Psycholinguistics*, 30: 225- 249.

Hung, Tony T. N. 2000. "Towards a Phonology of Hong Kong." In *Hong Kong English: Autonomy and Creativity*, ed. Kingsley Bolton, 119-140. Hong Kong: Hong Kong University Press.

Iverson, Paul, and Bronwen G. Evans. 2007. "Auditory training of English vowels for first language speakers of Spanish and German." In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS2007)*, ed. Jürgen Trouvain and William J. Barry, 1625-1628. Saarbrücken, Germany: Universität des Saarlandes.

Iverson, Paul, Melanie Pinet, and Bronwen G. Evans. 2012. "Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels." *Applied Psycholinguistics* 33(1): 145-160.

Jenkins, Jennifer. 2000. *The Phonology of English as an International Language: New Models, New Norms, New Goals.* Oxford, UK: Oxford University Press.

Kachru, Braj B. 1985. "Standards, codification and sociolinguistic realism: the English language in the outer circle." In English in the World: Teaching and Learning the Language and Literatures, ed. Randolph Quirk, Henry G. Widdowson, 11-30. Cambridge: Cambridge University Press.

Kim, Yoon Hyun, and Valarie Hazan. 2010. "Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates." In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, ed. Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel, and Małgorzata Kul, 251-256. Poznań, Poland: Adam Mickiewicz University.

Kuhl, Patricia K. 2000. “A new view of language acquisition.” In *Proceedings of the National Academy of Sciences of the United States of America* 97(22), 11850-11857.
Lambacher, Stephen G., William L. Martens, Kazuaki Kakehi, Chandrajith A. Marasinghe, and Garry Molholt. 2005. “The effects of identification training on the identification and production of American English vowels by native speakers of Japanese.” *Applied Psycholinguistics* 26(2): 227-247.
Leather, Jonathan. 1997. “Interrelation of perceptual and productive learning in the initial acquisition of second-language tone.” In *Second-Language Speech: Structure and Process*, ed. Allan R. James and Jonathan Leather, 75-101. Berlin: Mouton de Gruyter.
Lee, Jiyeon, Tyler K. Perrachione, Tasha M. Dees, and Patrick C. M. Wong. 2007. “Differential effects of stimulus variability and learners’ pre-existing pitch perception ability in lexical tone learning by native English speakers.” In *Proceedings of 16th International Congress of Phonetic Sciences (ICPhS2007),* ed. Jürgen Trouvain and William J. Barry, 1589-1592. Saarbrücken, Germany: Universität des Saarlandes.
Leung, Cynthia B., and Alejandro E. Brice. 2012. “An Analysis of Phonological Processes Involved in Spoken English of Hong Kong Primary Pre-service Teachers.” *Language Testing in Asia* 2(2): 42-60.
Lively, Scott E., John S. Logan, and David B. Pisoni. 1993. “Training Japanese listeners to identify English /ɹ/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories.” *Journal of the Acoustical Society of America* 94(3 Pt 1): 1242-1255.
Lively, Scott E., David B. Pisoni, Reiko Akahane-Yamada, Yoh’ichi Tohkura, and Tsuneo Yamada. 1994. “Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories.” *Journal of the Acoustical Society of America*, 96(4): 2076-2087.
Logan, John S., Scott E. Lively, and David B. Pisoni.1991. “Training Japanese listeners to identify English /ɹ/ and /l/: A first report.” *Journal of the Acoustical Society of America*, 89: 874-886.
MacKay, Ian R. A., Diane Meador, and James E. Flege. 2001. “The identification of English consonants by native speakers of Italian.” *Phonetica*, 58: 103-125.
Meng, Helen, Eric Zee, and Wai Sum Lee. 2007. “A Contrastive Phonetic Study between Cantonese and English to Predict Salient Mispronunciations by Cantonese Learners of English.” Technical

Report, SEEM2007-1500, Department of Sytems Engineering and Engineering Management. The Chinese University of Hong Kong.
Nishi, Kanae, and Diane Kewley-Port. 2005. “Training Japanese listeners to identify American English vowels.” *Journal of the Acoustical Society of America*, 117(4): 2401.
Nishi, Kanae, and Diane Kewley-Port. 2007. “Training Japanese listeners to perceive American English vowels: Influence of training sets.” *Journal of Speech, Language, and Hearing Research*, 50(6): 1496-1509.
Polka, Linda. 1991. “Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions.” *Journal of the Acoustical Society of America*, 89(6): 2961-2977.
Polka, Linda. 1992. “Characterizing the influence of native language experience on adult speech perception.” *Perception & Psychophysics*, 52(1): 37-52.
Pruitt, John S. 1995. “Perceptual training on Hindi dental and retroflex consonants by native English and Japanese speakers.” *Journal of the Acoustical Society of America*, 97(5): 3417-18.
Pruitt, John S., James J. Jenkins, and Winifred Strange. 2006. “Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese.” *Journal of the Acoustical Society of America*, 119(3): 1684-1696.
Rato, Anablea, Andréia Schurt Rauber, Denise Cristina Kluge, and Giane Rodrigues dos Santos. (this volumen). “Designing Speech Perception Tasks with TP”. In Mompean, J. & Fouz Gonzalez, J. (Eds.), *Investigating English Pronunciation: Current Trends and Directions.* Basingstoke and New York: Palgrave Macmillan[redacted]
Rochet, Bernard L. 1995. “Perception and production of second-language speech sounds by adults.” In *Speech perception and linguistic experience: Issues in cross-language speech research,* ed. Winifred Strange, 379-410. Timonium, MD: York Press.
Sewell, Andrew. 2009. “World Englishes, English as a Lingua Franca and the case of Hong Kong English.” *English Today*, 25: 37-43.
Slevc, L. Robert, and Akira Miyake. 2006. “Individual differences in second-language proficiency: Does musical ability matter?” *Psychological Science*, 17(8): 675- 681.
Sperbeck, Mieko, Winifred Strange, and Kikuyo Ito. 2005. “Training Japanese L2 learners to perceive difficult American vowel contrasts.” *Journal of the Acoustical Society of America*, 117(4): 2400.
Strange, Winifred. 1995. “Cross-language studies of speech perception: A historical review.” In *Speech perception and linguistic experience:*

Issues in cross-language research, ed. Winifred Strange, 3-45. Baltimore: York Press.

Strange, Winifred, and Sybilla Dittmann. 1984. “Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English.” *Perception and Psychophysics*, 36(2): 131-145.

Tajima, Keiichi, Hiroaki Kato, Amanda Rothwell, Reiko Akahane-Yamada, and Kevin G. Munhall. 2008. “Training English listeners to perceive phonemic length contrasts in Japanese.” *Journal of the Acoustical Society of America*, 123(1): 397-413.

Tsushima, Teruaki and Mayumi Hamada. 2005. “Relation between perception and production ability during a speech training course.” *Journal of the Acoustical Society of America*, 117(4): 2427.

Wang, Xinchun. 2002. “Training Mandarin and Cantonese Speakers to identify English vowel contrasts: long-term retention and effects on production” Ph.D. diss. Simon Fraser University. Available at: http://zimmer.csufresno.edu/~xinw/contents.pdf.

Wang, Xinchun, and Murray J. Munro. 1999. The perception of English tense-lax vowel by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS1999),* ed. John Ohala, 125-128. Berkely, CA: University of California.

Wong, Janice Wing Sze. 2012. “Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training.” In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology (SST 2012),* ed. Felicity Cox, Katherine Demuth, Susan Lin, Kelly Miles, Sallyanne Palethrope, Jason Shaw, Ivan Yuen, 37-40. Sydney, Australia: Macquarie University.

Wong, Janice Wing Sze. 2013. “The Effects of Training Diversity in Training the Perception and Production of English Vowels /ɪ/ and /i:/ by Cantonese ESL learners.” In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, ed. Frédéric Bimbot, Christophe Cerisara, Cécile Fougeron, Guillaume Gravier, Lori Lamel, François Pellegrino and Pascal Perrier, 2133-2117. Lyon, France: ISCA. Available at: http://www.isca-speech.org/archive/interspeech_2013/i13_2113.html.

Wong, Janice Wing Sze. 2014a. “The Effects of High and Low Variability Phonetic Training on the Perception and Production of English Vowels /e/-/æ/ by Cantonese ESL Learners with High and Low L2 Proficiency Levels.” In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014),* ed. Haizhou Li, Helen Meng, Bin Ma, Eng Siong Chng, Lei Xie, 524-528. Singapore. ISCA. Available at: http://www.isca-speech.org/archive/interspeech_2014/i14_0524.html.

Wong, Janice Wing Sze. 2014b. "Comparing the effects of perception and production training on the learning of English vowel contrast /e/ and /æ/ by Cantonese ESL learners." In *Contemporary Issues on Linguistics and Language (Linguistics and Language Conference Proceedings, LILA '14)*, ed. Farma Pinar Arslan, 317-326. Istanbul, Turkey: DAKAM Publishing.

Yamada, Tsuneo, Reiko Akahane-Yamada, and Winifred Strange. 1996. "Perceptual learning of Japanese mora syllables by native speakers of American English: An Analysis of Acquisition Processes of Speech Perception in Second Language Learning" In *Proceedings of the 3rd International Conference on Spoken Language Processing* (ICSLP'94), 2007-2010. Yokohama, Japan: ISCA. Available at: http://www.isca-speech.org/archive/icslp_1994/i94_2007.html.