CHAPTER IV

EFFECT OF PERCEPTUAL LEARNING ON PRODUCTION

4.1. INTRODUCTION

One of the goals of the current study is to examine the possibility of trainees’ perceptual learning of English vowel contrasts being transferred to production without undergoing any training in production. This chapter reports the results of two evaluation procedures - an intelligibility test and acoustic measurement - that are carried out to assess the effect of perceptual training on Mandarin and Cantonese subjects’ productions of the three pairs of English vowels under investigation.

First, in the intelligibility test, the trained and control groups’ productions of minimal pairs contrasting the three target vowel pairs at both pretest and post-test will be presented to a group of phonetically trained native English listeners for identification. The purpose of the identification task is to directly access any perceptible increase in intelligibility of trainees’ vowel productions at post-test. Because the trainees received perceptual training only, any significant increase in intelligibility scores in production at post-test seen with the trained but not the control group can be viewed as result of perceptual learning being transferred to production.

Second, vowel durations of the trained, the control, and a group of native English speakers’ productions of the three target vowel pairs will be measured. The purpose of the measurement is to examine whether the Mandarin and Cantonese subjects produced vowel inherent duration differences between /i/-/ɪ/, /u/-/ʊ/, and /e/-/æ/ contrasts appropriately and whether perceptual training affected the trainees’ productions of such duration differences.

The role of vowel duration in vowel perception has been explored in both L1
(Hillenbrand & Clark, 2000) and L2 perception studies (Bohn, 1995; Munro, 1993; Wang & Munro, 1999). Findings suggest that, in general, duration has a very limited effect on vowel identification for native speakers of English. In a recent study using synthesized vowels with four different duration values (original, neutral, short, and long), Hillenbrand & Clark (2000) found that the overwhelming majority of vowels with altered duration values were accurately identified as those with original duration values. However, some “modest” duration effects on the identification of some vowel pairs but not others were observed. For example, listeners’ identification of American English /æ/-/æ/-/æ/ and /æ/-/æ/ contrasts tended to be affected more by duration alternation (with modified duration values) than the /i/-/i/, /u/-/u/, and /i/-/e/-/e/ contrasts. While there were none or very few shortened /i/ and /u/ heard as /l/ and /u/, shortened /æ/ tokens were often heard as /æ/. Also, /æ/, /a/ and /ʌ/ were misidentified as one another when the duration values were altered.

On the other hand, studies of L2 vowel perception have shown that non-native speakers often respond to duration cues in identifying a synthesized English /i/ and /h/ continuum (Bohn, 1995; Wang & Munro, 1999). The use of duration cues to distinguish non-native vowel contrasts has been observed not only among those whose L1 system contrasts long and short vowels such as Arabic (Munro, 1993), but also with listeners whose L1 does not make the duration contrasts, such as Spanish and Mandarin (Bohn, 1995; Wang & Munro, 1999).

In production, the picture is quite different. Despite the fact that duration plays a very limited role in vowel perception in English, studies of vowel production have shown that

---

1 The reported modest effect of duration in identifying American English is /æ/-/æ/-/æ/ and /æ/-/æ/ contrasts may not apply to Canadian English because of dialectal differences.
native speakers of English generally produce durational differences in vowel pairs such as /i/-/ɪ/, /u/-/ʊ/, /e/-/æ/, and so on (Crystal & House, 1988; Hillenbrand, Getty, Clark, & Wheeler, 1995; Wang, 1997). Although the absolute vowel duration values vary considerably as a result of differences in speaking conditions such as speaking rate, forms (connected and citation), different stress conditions, pauses and boundaries, the inherent duration differences between long vowels /i e u o æ ø/ and short vowels /ɪ e u ø/ remain consistent in American English (Crystal & House, 1988; Hillenbrand & Clark, 2000).

In cross-linguistic vowel production studies, the findings have been mixed. For example, Munro (1993) found that native Arabic speakers exaggerated the duration differences between English tense and lax vowel pairs. Wang (1997), however, found that native Mandarin speakers produced a disproportionate duration difference on the English /i/-/ɪ/ contrast but not on the /u/-/ʊ/ and /e/-/æ/ contrasts. In a follow-up study using synthesized vowels, Wang & Munro (1999) found that Mandarin speakers responded heavily to duration cues for the /i/-/ɪ/ contrast only when they identified the English /i/-/ɪ/ and /u/-/ʊ/ continua. No parallel production test data were available.

In addition to the exploration of the effect of training on L2 vowel perception and production, the current study, which involves three pairs of English vowel contrasts (an additional /e/-/æ/ pair) in both perception and production, will examine how the Mandarin and Cantonese speakers relate their perception of intrinsic vowel duration differences (that are typically represented by these three contrasts) to their productions. In Chapter 3, the results of perceptual tests suggest that both the trained and the control groups responded to duration cues for the /i/-/ɪ/ contrast. In particular, six of the 16 trainees appeared to rely heavily on duration cues when identifying this particular pair. They systematically identified
the longer tokens as /i/s and short ones as /ɪ/s. However, for the /u/-/ʊ/ and /ɛ/-/æ/ contrasts, the use of duration cues was not common. For the few participants who actually responded to duration cues for the /u/-/ʊ/ and /ɛ/-/æ/ contrasts, the use of such temporal cues was not consistent, as longer tokens were sometimes identified as /ʊ/ and /ɛ/ rather than /u/ and /æ/.

As reported earlier, training effectively shifted trainees’ attention from duration to spectral cues at post-test, and the results of perceptual learning were retained in the 3-month test. It would be interesting to explore whether the participants in both groups produced the vowel duration differences of the three target vowel pairs accordingly.

Based on the results of the Mandarin and Cantonese participants’ performance in the perception test, it is predicted that both the trained and control groups may produce duration differences between the /i/-/ɪ/ contrast but not between /u/-/ʊ/ and /ɛ/-/æ/ contrasts at pretest. This is because some trainees perceptually responded to the duration cues to the /i/-/ɪ/ contrast only at the pretest. It is also possible that the trainees who responded to duration cues in identifying the synthesized heed/hid contrast may exaggerate the duration differences between /i/ and /ɪ/ in production at pretest. If these predictions are true, perceptual training may lead to reduced duration differences in productions between /i/ and /ɪ/, but not between /u/ and /ʊ/ nor between /ɛ/ and /æ/. The reduced sensitivity to duration cues for the /i/-/ɪ/ contrast through training might have an effect on their production.

4.2. INTELLIGIBILITY TEST

4.2.1. Method

4.2.1.1. Listeners

The listeners were four phonetically-trained native Canadian English speakers born and raised in Western Canada. They were graduate and senior undergraduate linguistics
majors who had taken advanced phonetics courses and who were all familiar with the
International Phonetic Alphabet (IPA). They all passed a pure tone hearing screen (20 dB,
250-4000Hz) before the experiment and all were paid for their participation.

4.2.1.2. Stimulus Preparation

In the pretest and post-test phases, as reported in Chapter 3, the control and trained
participants provided production data through two reading tasks. The recordings were made
immediately before the perception tasks at both pretest and post-test. The participants read a
list of words with two different elicitation procedures. The first reading task consisted of six
minimal pairs containing the target vowels in the sentence frame “Now I say ___”. A 13th
sentence with the same frame was added at the end to reduce the likelihood of a list effect.
The second task consisted of six answers to the six corresponding questions. Each answer
contained one or two target words that were also minimal pairs contrasting the vowels under
investigation. The reading list for both task 1 (Part A) and task 2 (Part B) is presented in
Table 4-1.

During the recording phase, the reading list (containing both Part A and Part B) along
with a written instruction about the reading and recording was presented to each participant
for brief preparation. The experimenter then explained the tasks to the speaker and clarified
any questions. For the first reading task, the target words were elicited with a “delayed
repetition” procedure (Flege, Mackay, & Meador, 1999; Munro, Wang, & Li, 2001) in which
both the printed form of the reading list and the aural prompt were provided.

The aural prompt for the first reading task (Part A) consisted of 13 recorded sentences
produced by a male native speaker of English. Each sentence contained a target word with
the frame “The next word is ____”. The target words in this frame matched those printed out
in the reading list to be read by the subject in a different sentence frame “Now I say ____”.
During the recording phase, the oral prompt (the recorded sentences) was presented in such a way that after each sentence was played from a tape recorder, there was a pause long enough for the subject to read the corresponding sentence (with the same target word in a different frame) from the reading list. Therefore, in addition to the differences in the sentence frame in which the target words were read, there was a delay in the repetition of the aural prompt. This elicitation method is preferred to simply reading for the reason that it may help the reader to avoid unnecessary orthographic confusion or “reading pronunciation” that underestimates a speaker’s real ability in production (Flege et al., 1999; Munro et al., 2001).

An “interrupted repetition” elicitation method in the form of answers to the questions was introduced in the second reading task. The target words, minimal pairs that contained the three pairs of vowels, were embedded in answers to questions that were presented aurally with visual prompts (see Part B of the reading list). For example, the reader’s answer “do you have a gas cooker in your kitchen?” containing the target words “gas” and “cooker” was elicited by the question, “what did he ask?”. The speakers had only the answers printed out in the reading list for them, but the questions were presented aurally via recordings produced by a male and a female speaker. In preparing the aural prompt for the second reading task, the male speaker began by asking the question to which the female speaker provided the corresponding answer immediately. The question and answer by the two native English speakers were recorded in sequence. Then, the male speaker asked the same question again for the reader to answer. The recording was made in such a way that the interval following the question was sufficient for the subject to produce the answers, but no extra time was allocated for thinking. It was also believed that the interruption by the recorded questions
would prevent the subject from direct repetition of the target words that occurred in the earlier answers produced by the female speaker. Also, unlike the previous task with the fixed sentence frame for each target word, in the second reading task, the carrier sentences for the target words in the form of answers to questions were all different. The advantage of embedding the target words in sentences with different frames was that the speakers did not know which were the words of interest. Therefore, the target words produced with this elicitation method added at least some variety in the speech data.

The recordings were made with a Macintosh computer using Sound Edit 16 software with 22.05 kHz sampling rate and 16-bit resolution. The speaker read the sentences into an MB quart K800 head-worn microphone connected to a Genexxa sound mixer that was connected to the computer. All productions were saved as waveforms for later analysis.

The target words were carefully separated from the sentence frames using waveform editing techniques, normalized for peak amplitude, and saved as audio files for playback. The stimuli used for assessment were nine minimal pairs containing the three target vowel contrasts from both pretest and post-test. Six pairs (two pairs per vowel contrast) were from the first reading task that was elicited in the sentence frame. The other three pairs (one pair per vowel contrast) were obtained from the answers to questions. Two of the minimal pairs were disyllabic words (beaten/bitten, cattle/kettle) that never occurred in training. The rest were all monosyllabic CVC words. The total number of tokens used for identification was 1512 (18 words × 2 productions × 2 tests × 21 speakers). They were presented to the listeners in seven listening sessions, each of which consisted of all the minimal pairs produced by three subjects. Two native English speakers’ productions of the same minimal pairs were used in the seven listening sessions in such a way that a complete set (one production, 18
stimuli) of the same minimal pairs by one native speaker of English was included in each listening session, making the total number of 234 stimuli per session (3 speakers × 18 words × 2 productions × 2 tests +18 native stimuli).

4.2.1.3. Procedure

Individual listening sessions were held in a sound-treated room using a Macintosh computer. The listeners identified all three vowel contrasts in each session, making it a six-alternative forced choice task. The labels “/i/ /I/ /u/ /U/ /eh/ and /ae/” were used for the six target vowels /i ɪ u ʊ æ/ respectively. Each listener heard the stimuli through a pair of headphones and identified the stimuli trial by trial by pressing one of the six buttons corresponding to the six vowels on the computer screen. As soon as the button was pressed, the next word was played. The listeners could hear each stimulus only once but had control over the time to make a decision by delaying their choices. Each listener completed the seven listening sessions on one day with breaks between sessions. The identification tasks took about 2.5 hours for each listener.

4.2.2. Results

4.2.2.1. Group Data

To examine the reliability of the six-way forced choice task, the percentage correct identifications of the tokens produced by the two native speakers of English were analyzed first. The average correct identification score on the native English speakers’ minimal pairs

---

2This task is different from the tasks the native Mandarin and Cantonese participants performed in their perceptual training and tests in which one pair of vowels was included in each session. Native English listeners are not expected to have perceptual problems with these vowel contrasts and, therefore, need not identify the vowels one pair at a time.

3 The labels were close to phonetic symbols except “eh” and “ae”, which stand for /e/ and /æ/ respectively. This was because the software for playback did not accommodate IPA symbols. These symbols, rather than key words, were used because the listeners were phonetically trained and quickly gained familiarity with the symbols after brief trial sessions.
was 98% across all listening sessions. It therefore seemed reasonable to assume that the listeners had performed reliably. Identification scores on both the trained and the control groups’ productions were then calculated.

Figure 4-1 presents the mean percentage correct identification scores (pooled over the minimal pairs produced with two elicitation tasks)\(^4\) of the trained (top) and control (bottom) groups’ productions on all three target vowel contrasts in pretest and post-test as identified by four native speakers of English. On average, for the trained group, the percentage increase from pretest to post-test was 6% for the /ɪ/-/ɪ/ pair, 9% for the /ʌ/-/ʌ/ pair and 4% for the /ɛ/-/æ/ pair. With a modest increase in intelligibility scores of 4%-9% across three vowel contrasts, the trained group’s average production scores, that ranged between 71%-76% at post-test, were still far behind the two native English speakers’ production scores of 98%.

For the control group, at post-test, there was a 2% increase the /ɪ/-/ɪ/ pair but a 4% and 5% decrease for the /ʌ/-/ʌ/ and /ɛ/-/æ/ pairs respectively.

A two-way repeated measures ANOVA with groups (trained and control) as a between factor, test (pre and post) and vowel (three pairs) as within factors showed the main effect of group \([F(1,19) = .762, p = .3935]\), test \([F(1,19) = 1.050, p = .3184]\), the interaction of vowel × group \([F(2,2) = 1.344, p = .2730]\), vowel × test \([F(2,1) = .654, p = .5255]\), and vowel × test × group \([F(2,2,1) = .651, p = .5270]\) all failed to reach significance (at \(p < .05\)). The effect of vowel \([F(2,19) = 4.746, p = .0145]\) and the test × group interaction \([F(2,2) = 5.116, p = .0356]\) were significant. However, subsequent tests of simple effects did not establish significant differences (at \(p < .05\)) between the groups at pretest or at post-test. Post-

\(^4\) Data from two reading tasks with different elicitation methods were not analyzed separately in this study because the purpose was to get a range of speaking data rather than comparing the effect of speaking condition on intelligibility.
hoc Tukey tests showed the effect of vowel was due to differences between /i/-/u/ and /u/-/o/,
and between /i/-/u/ and /e/-/æ/ pairs. Both groups produced a more intelligible /i/-/u/ contrast
than /u/-/o/ and /e/-/æ/ contrasts. The effect of vowel was in agreement with the results of the
perception tests in which they identified the /i/-/u/ contrast significantly better than the /u/-/o/
and /e/-/æ/ pairs.

The results of the ANOVA gave no indication that the trained group in general
outperformed the control group in producing the three target vowel contrasts in terms of
intelligibility at post-test. The trained group’s significant improvement in perception was not
accompanied by significantly better performance in production. Therefore, there was no
evidence that perceptual learning through training was transferred to the production mode in
terms of intelligibility. Yet, some non-significant increases in production scores on each of
the three target vowel contrasts were observed with the trained but not the control group. For
this reason, each individual trainee’s production scores were analyzed to examine how the
changes in scores ranged among the trainees and how individual performance in production
was related to perception of the target vowel contrasts.

4.2.2.2. Individual Differences

Each participant’s production scores on the three target vowel pairs at pretest and
post-test, and the differences between the two tests are presented in Table 4-2. Considerable
individual differences were observed among the trainees but, in general, three different
patterns of performance emerged. Six speakers, (C01, M04, M08, M14, M16, and M20)
showed an overall increase in intelligibility scores that ranged between 11%–22% averaged
across three pairs from pretest to post-test. Four speakers (M05, M09, M12, and M15) had a
modest increase of 5%–8%. The remaining four trainees showed no noticeable change (C02,
M03, M07 and M13) as their post-test scores were -1% to +3% different from the pretest. However, two trainees had decreases in production scores from pretest to post-test by 9% (C03) and 6% (M17).

Trainees’ productions across the three target vowel contrasts also varied in terms of increased % ID scores at post-test. For example, of the six speakers who demonstrated the greatest overall increase in intelligibility scores, M14 had an increase of 33% on the /u/-/o/ pair but no change in scores on /i/-/I/ and /e/-/æ/ at all. C01 and M20 both seemed to have gained noticeably on the /i/-/I/ and /u/-/o/ pairs but not on /e/-/æ/. Only M04 and M16 appeared to have a noteworthy increase in intelligibility scores on all three pairs. Still, the level of gains among the three pairs differed noticeably. However, both M09 and M15, who showed an average modest increase in scores across the three vowel contrasts had substantial amount of increase (25% and 10%) in production scores on the /e/-/æ/ pair. Even trainees who did not show overall gains on the average, such as M03 and C02, had 10% -13% increases on some individual vowel pairs. On the other hand, there were also those who had noticeable decreases in scores on individual vowel pairs. For example, for each vowel contrast, there were 1-2 speakers who had a decrease of 10%-13% in scores.

Detailed analysis of individual data among the control subjects showed that C05 had a 17% increase in production scores on the front vowel pair. However, across the three target vowel pairs, this was the only case in which a control subject showed an increase in scores that was above 10%. In contrast, far more individual trainees had similar noticeable increases in scores. Compared with the trained subjects’ individual data, the number of control subjects who had such a noticeable increase in scores at post-test was out of proportion.

In summary, considerable individual differences were found with the trainees in terms
of increased or decreased scores on each of the three vowel pairs. As the ANOVA test already showed that the trained group in general did not outperform the control group in terms of increased intelligibility in their productions of the three target vowel contrasts, the individual differences observed among the trainees may simply reflect a range of variability. The descriptive analyses of such individual data often raise questions rather than provide answers. However, on each vowel contrast, there were 6-7 trainees who showed noticeable increases in production scores at post-test. Proportionally, the control subjects did not appear to match these increases in scores seen with the individual trainees. Some interesting changes may have taken place in the trainees’ productions after the long-term training. It would be interesting to examine those trainees’ productions in the 3-month test.

4.2.3. Comparison between Perception and Production

To compare how the trainees’ performance in one mode is related to the other, each trainee’s perception and production scores on the three target vowel pairs in both pretest and post-test are summarized in Table 4-3 through Table 4-5. Visual observations show that, in general, the trainees demonstrated a greater increase in scores in perception than in production. There were also differences in performance and degree of learning across three vowel contrasts.

For the /i/-/u/ pair (see Table 4-3), the pretest scores and the degree of increases were generally higher in perception than in production. Nine of the 16 trainees’ perception scores reached 100% in the post-test, although the amount of increase for these trainees in the perception mode was apparently affected by a ceiling effect. The remaining 10 trainees’ post-test perception scores all increased to above 92%. In production, the pretest scores were
much lower than those of the perception test and thus, had more room for improvement. Yet, the size of gain in production was much smaller than that seen for perception.

When individual performance in perception and production was compared, three different patterns emerged. First, as seen from Table 4-3, three trainees (C01, M05, and M12) showed noticeable but different increases in intelligibility scores in both perception and production. Four others (M03, M04, M16, and M20) also had a noticeable increase with the range of 10% - 27% in production although their size of gain in perception was not as large due to an apparent ceiling effect. Second, four trainees (C02, C03, M09, M17) showed an increase in perception scores that exceeded 20% but no increase in production at all from pretest to post-test. Also, two trainees (M13 and M14) showed a modest increase in scores in perception (due to apparent ceiling effect) but no gains in production.

With these differences in the sizes of increase and decrease in the scores in the two modes, each individual’s learning pattern appeared to be different. For example, M05 and M09 had about the same level of performance in both perception and production at pretest and their gains in the perception domain were also comparable at post-test. However, the level of production performance was widely different at post-test. On the other hand, M03 had no increase in perception scores as her pretest scores already reached 100% at pretest. Her production scores went from 85% to 96% at post-test. M13 had a high level of performance in perception at pretest and also a relatively high production score of 85% that was the same as M03. However, his production score at post-test was unchanged. C01 and M16 had a noticeable increase in scores in both perception and production although their pretest performance and level of gains differed in both modes.

For the /u/-/o/ pair, differences among trainees in terms of degree of change in
perception and production scores were also common. In general, while pretest scores in both perception and production were low, a more substantial increase in scores was found in perception than in production. Eight of the 16 trainees had a pretest to post-test increase of scores between the range of 27% -100% in perception. (Some of them might have confused the key words in the perception tasks as suggested by a 100% shift in scores). In the production domain, six trainees (C01, M04, M08, M14, M16 and M20) showed a noticeable increase in scores that ranged from 17% to 33%. Four (C01, M08, M14, and M16) had an increase in scores in both domains, although the level of gains was very different. Comparing the degree of gains in both domains, wide individual differences were found. For example, C02 and C03 had a 59% change in perception after training, yet, their production scores had a negative change. In contrast, M20, whose performance in perception on this vowel pair was 100% accurate but a 67% score in production, had a 17% increase in production at post-test.

For the /ɛ/-/æ/ pair, on average, although the trainees had almost the same level of performance in perception (68%) and production (67%) at pretest, the degree of gains was much greater in perception than in production in the post-test. Ten of the 16 trainees had a percentage change between 12%-38% from pretest to post-test in perception while only six had a 10% - 33% gain in production. Five trainees (C02, M04, M08, M09 and M15) had a noticeable increase in both domains although the level of gains varied for all of them. For example, for M08, who had exactly the same scores of 58% in perception and in production at pretest, the increase in perception (30%) greatly exceeded that in production (10%) at post-test. Similarly, another three trainees (C01, M17 and M20) also had noticeable increases in scores in perception at post-test. Their production scores did not change or even went down at post-test. In contrast, one trainee, M16, had no sign of increase in scores in perception at
all but had a 33% increase in her production scores in the post-test.

Summing up, individual trainees’ gains in perception and production varied greatly on all three target vowel pairs. First, there were trainees who had noteworthy increase in scores in perception but had no or very little gains in the production mode. On each of the three vowel contrasts, trainees who had increases in scores in perception outnumbered those who had noticeable gains in production. Such differences were expected because the participants received training in perception only. These differences were also in agreement with the general findings of significant improvement for the trained group in the perception but not in the production mode as indicated by the ANOVA tests. However, on each one of the three target vowel contrasts, there were trainees who had noticeable increases in scores in both perception and production, although, in general, the sizes of gains in both domains were not parallel. Nevertheless, in these cases, better perception appeared to be accompanied by better production.

4.2.4. Three-Month Test

As discussed earlier, overall, the statistical analyses showed that the trained group’s increase in production scores failed to reach statistical significance. There was no evidence of improvement in the production mode as a result of perceptual training. Yet, the examination of individual data showed that some trainees had noticeable increases in production scores at post-test. On each vowel contrast, there were 6-7 trainees who showed noticeable increases in production scores at post-test. (See Section 4.2.2. for detailed analyses).

To follow up whether these trainees retained these gains in production after the training, another intelligibility test was carried out three months later. Trainees who did not have noticeable gains in production scores at post-test were excluded from this test. The
criteria for selecting the trainees for the follow-up test were a 15% increase in scores from pretest to post-test for the /i/-/ɪ/ and /u/-/ʊ/ pairs and a 10% increase in scores for the /e/-/æ/ pair. The reason for choosing this standard was that a 10%-15% increase was noticeably greater than the average increase. Based on this standard, the number trainees included in the 3-month test were five, six, and six for the /i/-/ɪ/, /u/-/ʊ/, and the /e/-/æ/ pairs respectively.

4.2.4.1. Method

In the three-month test on perception reported in Chapter 3, the trainees performed the same two reading tasks that they did at pretest and post-test phases (See the above section for a detailed description of reading tasks and recording procedures.) The productions of the relevant participants were included in the intelligibility test. The same minimal pairs that were used in the post-test were selected from the speakers’ 3-month productions as test tokens. Three minimal pairs were used for each vowel contrast. The total number of test words included in the identification task was 204. There were 30 minimal pairs or 60 tokens for the /i/-/ɪ/ contrast (5 talkers × 3 pairs × 2 productions), 72 tokens for the /u/-/ʊ/ and /e/-/æ/ contrasts each (6 talkers × 3 pairs × 2 productions). The tokens were mixed and presented in two listening sessions, each of which covered all three vowel contrasts. In each test session, the same minimal pairs produced by a native English speaker were also included. The identification test was the same six-way forced-choice task as the previous intelligibility test and used the same six labels, /i/ /ɪ/ /u/ /ʊ/ /æ/ and /æ/ for the six target vowels /i ɪ u ʊ e æ/ respectively. Four phonetically trained native speaks of English, one of whom also performed the previous listening tasks, participated as listeners. The procedures of identification tasks were the same as in the post-test.
4.2.4.2. Results

The selected trainees’ production scores at 3-month test are presented in Table 4-6, along with their scores at pretest and post-test in both the perception and the production modes for comparison. Across all three vowel contrasts, there was a decrease in production scores in the 3-month test from the post-test. On average, the score for the /i/-/u/ contrast dropped to 64% from 79%. (i.e., to the level of the pretest.) For the /u/-/ʊ/ pair, it went down to 64% from 83%, only 4% higher than the pretest scores. The decline in scores from post-test to 3-month test was relatively less but still noticeable for the /ɛ/-/æ/ contrast, from 75% to 70%.

4.2.4.3. Production and Perception Comparison

As shown in Table 4-6, a general pattern of decline in production but not perception scores from post-test to 3-month test was observed. However, differences in degree of “gains” and “loss” were found in perception and production with three vowel contrasts. While the perception scores for the /i/-/ɪ/ and /u/-/ʊ/ remained almost the same as they were at post-test, the scores in production dropped to the pretest level. For the /ɛ/-/æ/ pair, the overall size of decrease in scores was similar, around 5% to 6%. Yet, the selected trainees’ scores on the /ɛ/-/æ/ pair were still lower in production than in perception.

Individuals differed in terms of size of loss from post-test to 3-month test. Three of the six subjects (M04, M15, and M16) had, in fact, retained their scores on the /ɛ/-/æ/ contrast in perception and production at 3-month test. None of the selected trainees had retained their production scores at the post-test phase for the /i/-/ɪ/ and /u/-/ʊ/ pairs at the 3-month test. This general pattern of decline in production scores at 3-month test was very different from what was observed at post-test. What appeared to be some noticeable increase
in production scores at post-test was disappeared almost completely. In contrast, their gains in post-test perception scores were generally retained.

4.2.5. Summary of Intelligibility Test

The trained group’s modest increase in production scores on all three vowel pairs at post-test was not significantly different from that of the control group. There was no evidence that learning in perception was transferred to production. This outcome was not totally unexpected because the subjects underwent training in perception only. The non-significant gains in production scores by the trained group may suggest that certain changes took place in their productions. Descriptive analysis of individual data showed that there were interesting individual differences in trainees’ performance. On each of the three vowel contrasts, a few trainees demonstrated a more noticeable increase in their production scores than others. However, even for those trainees who did have noticeable increases in their production scores at post-test, the results of the 3-month test showed that such increases disappeared almost completely, especially for the /i/-/ɪ/ and /u/-/ʊ/ contrasts three months after training was completed. In contrast, their increases in scores in perception were maintained at the 3-month test. A direct comparison of test scores in perception and production at post-test and 3-month test (selected subjects) showed that trainees’ performance in production did not match their perception. The findings suggest that training in the perception mode alone was not sufficient for improvement in production on the three pairs of vowels under investigation.

4.3. ACOUSTIC ANALYSIS: VOWEL DURATION

4.3.1. Method

The tokens selected for measurement of vowel duration were heed/hid, who’d/hood,
and head/had that were collected in both pretest and post-test from the first reading task. A detailed description of production data collection was given in the above section in the intelligibility test. The reason for choosing only these three pairs (more minimal pairs were used in the intelligibility test) was that they all occurred in the same phonetic environment /hVd/ and were elicited in the same sentence frame. Other minimal pairs were not included for the duration measurements because they either occur in different consonant environments across the three vowel contrasts, or were elicited from the answers to the questions with the “interrupted repetition” procedure and occur at different positions in the utterances. Different phonetic contexts and different positions in an utterance may have an effect on vowel duration.

A total of 24 words per speaker were included (3 vowel pairs × 2 productions × 2 tests). In addition to the 21 trained and control subjects’ productions, for comparison purposes, the same minimal pairs produced by 7 native English speakers (among them, six speakers’ productions were used as training and test tokens) were also included for measurement. The native speakers’ productions of these vowel pairs were elicited with the same sentence frame used in the reading task. Vowel duration was measured from waveforms using Sound Edit 16 software on a Macintosh computer.

4.3.2. Results

The mean durations of vowels /i ɪ u ʊ æ/ with standard deviation for both trained, control, and English groups are summarized in Table 4-7. Also presented in Table 4-7 are duration ratios of the three vowel pairs which were calculated by dividing the mean duration values (in milliseconds) of /i/, /u/ and /æ/ by the mean duration values of /ɪ/, /ʊ/ and /æ/ respectively. The higher the values of these ratios are, the greater the differences between the
two vowels of the pairs are. Therefore, the duration ratios reflect the vowel duration
differences between the tense and lax vowels of the three pairs under investigation.

As seen in Table 4-7, although the absolute duration values on each vowel differed
across groups, the duration ratio for the /i/-/i/ pair was 1.4 for all three groups at pretest. On
average, the control group produced somewhat longer /i/ and /i/ than the English and trained
groups. As expected, both the control and trained Mandarin and Cantonese speakers
produced longer /i/ than /i/, but the ratio suggests that the difference was comparable to the
native English group’s productions. However, for the /u/-/u/ pair, the duration ratios were 1.1
for both the trained and control groups. In contrast, the native English group had a much
higher duration ratio (1.6) for the /u/-/u/ pair. For the /æ/-/æ/ pair, the duration ratios for the
trained and control groups were 1.1 and 1.2 respectively, also noticeably lower than that of
the native English speaker group of 1.4.

At post-test, the trained group showed a slight increase in duration ratios on all three
vowel pairs. The change was visually summarized in Figure 4-2. However, the same modest
increase in duration ratios was not observed with the control group. A two-way repeated
measures ANOVA with group (trained and control) as a between factor, vowel contrast (3
levels) and test (pre and post) as two within factors revealed the main effects of group
[F(1,19) = 1.296, p = .2690], and test [F(1,19) = 2.563, p = .1259] were not significant. The
effect of vowel was significant [F(2,19) = 7.405, p = .0025]. However, the test × vowel
contrast interaction [F(2,38) = .337, p = .716], group × vowel interaction [F(92,380 = .322, p
= .7270], and group × vowel × test interaction [F(2,38) = .218, p = .8052] all failed to reach
significance (at p < .05). Post hoc Tukey tests (at p < .05) on vowel contrast established
differences between the /i/-/i/ and the /u/-/u/ pairs, and between the /i/-/i/ and /æ/-/æ/ pairs.
Both the trained and the control groups produced a significantly higher duration ratio for /i/-
/u/ than for the /u/-/u/ and /e/-/æ/ pairs. The trained group’s increased duration ratio for the
/u/-/u/ and /e/-/æ/ pairs at post-test was not significant but showed a tendency toward more
native-like duration ratios in their productions.

Considerable individual differences in vowel duration ratios among all three groups
were observed. The front vowel pair was chosen to analyze these individual differences and
each individual subject’s vowel duration values of the /i/-/u/ productions at pretest and post-
test are summarized in Table 4-8. A particularly interesting point for the analysis of
individual speaker data of this pair was to track the trainees who relied heavily on the
duration cues for the identification of the synthesized /i/-/u/ tokens in the perceptual test. As
reported in Chapter 3, six trainees (C01, M04, M07, M09, M16, and M17) identified the
longer tokens as /i/s and short ones as /u/s, ignoring the spectral differences. It is possible that
these speakers may produce greater duration differences on the /i/-/u/ pair. However, as seen
in Table 4-8, these trainees did not, in fact, exaggerate the duration differences between the
/i/-/u/ pair. Only one of the six subjects produced a somewhat higher duration ratio (1.8) than
the group average of 1.4. Two subjects had duration ratios close to group mean and three
actually had a low duration ratio of 1.1-1.2, lower than the group mean. Compared to other
trainees who did not perceptually rely on duration cues for /i/-/u/ identification, these
duration-oriented subjects did not produce noticeably longer /i/s or shorter /u/s than the rest of
the group. Interestingly, at post-test, the majority of the trainees, including these six duration
oriented subjects, slightly increased the duration differences for the /i/-/u/ pair.

In summary, although both the trained and the control group produced the same
duration ratio for the /i/-/u/ contrast as the English group, both groups failed to produce
native-like duration differences between /u/-/o/ and /ɛ/-/æ/ contrasts at pretest. The two non-native groups did not exaggerate the duration difference between /i/ and /ɪ/ in their productions as was expected based on their sensitivity to the duration differences on this pair compared with the other two pairs. In particular, an examination of individual data showed that the six duration-oriented subjects in the perceptual test did not produce a greater duration difference for /i/-/ɪ/ contrast than average. At post-test, the trained group had a non-significant increase in duration ratios, especially for the /u/-/o/ and /ɛ/-/æ/ contrast that approximated the native-English speaker group.

4.3.3. Discussion

Despite the fact that duration plays a very limited role in vowel perception in English, the results of acoustic measurement showed that, on average, the native English speaker group produced clear duration differences between the tense and lax vowels on all three vowel pairs under investigation. In contrast, both the trained and control groups showed an inconsistent pattern in their productions of vowel duration across the three vowel pairs.

First, both the trained and the control groups produced normal but not disproportional duration differences between the /i/ and /ɪ/. Moreover, the six duration-oriented trainees did not appear to exaggerate duration differences between the /i/ and /ɪ/ in their productions either, although they responded heavily to duration cues in identifying the synthesized heed/hid contrast. The finding was not in agreement with an earlier study in which a group of 15 native Mandarin speakers produced a much greater duration difference between /i/ and /ɪ/ (Wang, 1997).

On the other hand, both the trained and the control groups produced smaller duration differences between /u/-/o/ and /ɛ/-/æ/ at pretest as compared with the native English group.
The reason that Mandarin and Cantonese speakers failed to produce a proper duration difference between these two pairs might be related to the fact that they were not relying on duration cues for contrasting these two vowel pairs. As reported in Chapter 3, the trained and control subjects did not show a systematic response to duration cues to the /u/-/ʊ/ and /ɛ/-/æ/ contrasts in their identification tasks. It is possible that Mandarin and Cantonese speakers need to perceive the duration differences between the vowels in order to produce the differences. In this respect, the reduced duration differences between /u/ and /ʊ/ and between /ɛ/ and /æ/ appeared to be in agreement with their perceptual patterns. If the subjects did not need to produce an exaggerated duration difference to bring out the /i/-/ɪ/ duration contrast that many of them were sensitive to in their perception, then they probably did not need to produce any duration differences at all for the /u/-/ʊ/ and /ɛ/-/æ/ duration contrasts that most of the subjects were not even perceptually sensitive to.

Perceptual training did not seem to have much impact on trainees’ productions in terms of vowel duration differences. As the ANOVA test showed, the trained group did not show significant changes in duration ratios on the three target vowel contrasts as compared with the control group at post-test. The trainees did not need to reduce duration ratios for the /u/-/ʊ/ and /ɛ/-/æ/ pairs because they did not produce any significant duration differences on these two pairs at pretest in the first place. As the training shifted listeners’ attention away from duration to spectral differences for the /i/-/ɪ/ contrast, it was expected that the trained group might reduce the duration differences between /i/ and /ɪ/ as a result of training. Yet, training did not appear to have helped to reduce the duration ratios on the /i/-/ɪ/ contrast either. The trained group continued to produce a normal duration difference for the /i/-/ɪ/ pair at post-test.
The interesting thing to point out was that the trained but not the control group had some slight increase in the duration differences between the long and the short vowels for the /u/-/ʊ/ and /ɛ/-/æ/ pairs in their productions at post-test. Although statistically not significant, the increase in duration ratios for the tense and lax vowel duration differences moved toward the ratio seen in the native English productions. More importantly, such changes led to more consistent duration differences between the long and short vowels across three pairs in their productions.

Another interesting observation made with the current data was the differences between L1 and L2 speakers in using duration cues in perception and production of English vowel contrasts. Earlier studies (Hillenbrand & Clark, 2000; Wang & Munro 1999) have concluded that, for native English speakers, their awareness of the inherent vowel duration differences generally does not influence their identification of vowel contrasts between long and short vowels (the tense-lax distinction). However, native English speakers often produce the inherent duration differences between the tense and lax vowel contrasts (Munro, 1993; Wang, 1997). The current data provided further evidence for such differences in vowel durations between perception and production.

For Mandarin speakers, the production of vowel duration differences on English tense and lax vowel pairs does not appear to be consistent. In the current study, both the trained and the control groups produced a significantly larger duration difference between /i/ and /ɪ/ than between /u/ and /ʊ/ and between /ɛ/ and /æ/. In an earlier study, Wang (1997) also found that native Mandarin speakers exaggerated the duration differences between /i/ and /ɪ/ (with a duration ratio of 2) but not between /u/ and /ʊ/ (with a duration ratio of 1.3) nor between /ɛ/ and /æ/ (with a duration ratio of 1.2). The findings suggest that differences in duration ratios
in the Mandarin and Cantonese speakers’ production of English tense and lax vowel pairs might be related to their differences in sensitivity to duration cues in identifying these vowel contrasts.

Summing up, this chapter reported the two evaluation procedures - an intelligibility test and acoustic measurement of vowel durations. Results showed that there was no evidence indicating any significant improvement in the trained group’s productions at post-test. Referring back to the research question raised for the production study at the beginning of this chapter, the trained group’s production scores of the three target vowel contrasts increased modestly at post-test but the increase in intelligibility scores was non-significant as compared with that of the control group. Similarly, perceptual training did not appear to have a significant effect on vowel durations in production. As the trainees did not produce any exaggerated duration differences between the /i/-/ɪ/ pair and did not even produce the normal duration differences between /u/-/ʊ/ and /e/-/æ/ contrasts at pretest, they did not (and did not need to) reduce the duration differences between these tense and lax vowels. Although the trained group produced a more native-like duration difference between the /u/-/ʊ/ and /e/-/æ/ contrasts at post-test, this change in duration ratios did not reach significance either when compared with that of the control group.