CHAPTER V

GENERAL DISCUSSION

5.1. AN OVERVIEW OF THE CURRENT STUDY

The main goals of the current study were to examine the effects of long-term training on the perceptual learning of L2 vowel contrasts and to explore whether any observed improvement in perception was retained and transferred to production.

Overall, the results presented in Chapter 3 suggest that perceptual training on three English vowel contrasts (/i/-/u/, /u/-/u/, and /e/-/æ/) was effective in modifying Mandarin and Cantonese speakers’ perceptual patterns for the target vowel contrasts. First, trainees who either relied heavily on the temporal cues or did not make effective use of the spectral cues in identifying the synthesized stimuli in the pretest attended to spectral cues in the post-test. Second, the trained group’s perceptual identification scores on natural tokens increased significantly (by 14% for the /i/-/u/ pair, 32% for the /u/-/u/ pair, and 16% for the /e/-/æ/ pair) from pretest to post-test. Third, in the test of generalization, the trainees identified the new words produced by new talkers as well as they identified familiar words and new words produced by familiar talkers. Therefore, perceptual learning was generalized to new talkers and new words. Fourth, the trained group’s perceptual performance on both the synthesized and the natural tokens on all three pairs of vowel contrasts at the 3-month test was comparable with their performance at the post-test. Perceptual learning was retained three months after training was completed.

The production test described in Chapter 4 explored the possibility of transfer of perceptual learning to production. The assessment of possible improvement in production
was carried out using two techniques: an intelligibility test and an examination of acoustic measurements of vowel durations of the trained and control subjects’ productions.

The results of the intelligibility test showed that the trained group in general did not outperform the control group in producing the three target vowel contrasts in terms of increased intelligibility scores at post-test. The trainees’ improvement in perception was not accompanied by significantly better performance in production. Therefore, there is no evidence that perceptual learning through training was transferred to production, although on each vowel contrast, there were 6-7 trainees who showed noticeable increases in production scores at post-test. However, any apparent increases in production scores at the post-test phase were lost almost completely in a follow-up test that tracked these trainees’ productions three months after the training was completed. In contrast, the 3-month test on perception showed that the trainees’ improvement in perception of these vowel pairs was retained.

The results of acoustic measurements of vowel duration differences between the vowel pairs yielded no evidence that perceptual training influenced the trainees’ productions significantly in terms of vowel duration. The trained and control groups did not produce the inherent duration differences (typical of native English speakers) between /u/ and /ʊ/ or between /ɛ/ and /æ/ (as shown by the low duration ratios) at pretest. Their failure to produce native-like duration differences between these two vowel pairs might have been related to their perceptual representations of the target phones. Recall that in the perceptual tests, the subjects did not rely on duration cues for the /u/-/ʊ/ and /ɛ/-/æ/ contrasts as they did with the /ɪ/-/ɪ/ pair. It might be expected that the trainees, especially those who relied on duration cues for the /i/-/ɪ/ contrast, would produce a disproportional duration difference between /i/
and /u/. However, contrary to expectation, the trained and control groups did not exaggerate, but produced normal (as compared with the native English group) duration differences between /i/ and /u/ at pretest. It might also be expected that the trainees would reduce the duration ratios for the vowel pairs in their productions at post-test because training focused their attention away from temporal cues. However, overall, there was no evidence that training caused a reduction in the duration ratio for the /i/-/u/ pair or increased duration ratios for the /u/-/o/ and /e/-/æ/ pairs. After all, the trainees did not need to reduce the duration ratios for the three vowel contrasts, as they actually produced a native-like duration ratio for the /i/-/u/ pair and low duration ratios for the /u/-/o/, and /e/-/æ/ pairs in the first place.

Interestingly, some trainees exhibited increases in the duration ratios for the /u/-/o/ and /e/- /æ/ pairs at post-test. Though statistically not significant, the increase in duration differences between these vowel pairs was in the right direction.

5.2. DIFFERENCES IN LEARNING BETWEEN VOWEL CONTRASTS

The results of the perception and production tests showed that there was an effect of vowel contrast at both pretest and post-test. The front vowel pair /i/-/u/ was consistently better identified and better produced than the /u/-/o/ and /e/-/æ/ pairs. It is not immediately clear that the causes of the differences in perceptual patterns and asymmetrical development in learning across the vowel pairs can simply be explained by the differences in L1 and L2 vowel systems. For example, although Mandarin has both /i/ and /u/ but lacks /i/ and /u/ categories in its vowel system and Cantonese has both lowered allophonic [I] and [U] in closed syllables, the results of the pretest suggest that Mandarin and Cantonese speakers
perceived the synthesized /i/-/u/ continuum as two different categories (often as long and short vowels) but probably perceived a single category for the /i/-/u/ pair. Several possible explanations for the observed asymmetry in learning across three vowel pairs will be discussed in this section.

5.2.1. Perceived Category Differences and L2 Perception Models

One cause for the differences in learning might be related to the differences in listeners’ perceived category distinctions across the three target vowel pairs at pretest. These differences will be discussed with respect to the current L2 speech perception models, i.e., Best’s Perceptual Assimilation Model (PAM) and Flege’s Speech Learning Model (SLM). More detailed accounts of these models were presented in Chapter 1. A brief summary of the PAM and SLM is provided here.

The PAM model predicts four possible assimilation patterns of two L2 sounds to the listeners’ L1 system: Two Category (TC) assimilation, Single Category (SC) assimilation, Category Goodness (CG) assimilation, and Non-Assimilation (NA). The model also predicts that the level of difficulty with discrimination is related to the type of assimilation. For example, it is predicted that the Single Category assimilation in which two L2 phones are assimilated to a single L1 category would be more difficult for listeners than Two-Category and Category Goodness assimilations. Flege’s Speech Learning Model states that L2 categories can be established when the differences between L1 and L2 sounds or between two different L2 sounds are perceived. The SLM also predicts that listeners’ productions of L2 contrasts will eventually correspond to their perceptual representations of the non-native contrasts. The SLM is a dynamic model that takes into consideration improvement over time and relates perception to production in learning.
For a better understanding of these assimilation models, a brief description of L1 and L2 segmental categories is necessary. Neither Mandarin nor Cantonese contrasts the three English vowel pairs under training in the current study. However, Cantonese does have [i] and [u] surface forms in certain phonetic environments and Mandarin has an allophonic [ɛ] form that is realized only after an onglide. A detailed description of Mandarin and Cantonese vowel systems in comparison to the English vowel system was presented in Chapter 2.

The current findings suggest that the participants in this study had different assimilation patterns for the three target vowel pairs. As demonstrated by their random patterns in responding to both the spectral and the duration cues in the synthesized continuum, the majority of the participants did not have a clear category distinction for the /ɛ/-/æ/ pair. Applying the PAM Model, a Single Category assimilation may be proposed for the /ɛ/-/æ/ pair. Although the Mandarin system does not have a clear /ɛ/ or /æ/ category as monophthongs, it does have a surface [ɛ] which occurs after onglides /i/ and /u/ such as in [ie] “leaf” and [tɛ] “moon”. It is possible that the Mandarin subjects assimilated both /ɛ/ and /æ/ to the Mandarin [ɛ] element as in [ieɛ] and [tɛɛ].

Given the fact that both English /i/-/ɪ/ and /u/-/ʊ/ contrasts differ in terms of duration and spectral properties in a similar way and that Mandarin lacks both /i/ and /ʊ/ but has /i/ and /u/ categories and Cantonese has /i/ and /ʊ/ in closed syllables, it is possible that Mandarin and Cantonese listeners may have assimilated the /i/ and /ʊ/ categories to the corresponding /i/ and /u/, either as Single Category (SC) or as Category Goodness (CG) assimilation. However, the results of the perceptual tests showed that the participants had
different assimilation patterns for the two vowel contrasts. First, as discussed earlier, participants who had problems with the /i/-/u/ pair appeared to identify the /i/-/u/ pair as two different categories, although many attended to the duration cues for the contrast. In contrast, those who had problems with the /u/-/u/ pair did not clearly respond to either spectral or duration differences. This random perceptual pattern suggests that at least some subjects did not have a clear two-category perception for the /u/-/u/ contrast. Therefore, Mandarin and Cantonese listeners’ inability to distinguish the /u/-/u/ pair may be due to Single Category assimilation. At the same time, one might propose that the participants’ perception of /i/-/i/ pair is a case of Category Goodness assimilation. However, the problem is that Mandarin and Cantonese subjects who identified the /i/-/i/ contrast using duration cues did not identify the phonetically short /i/ as /i/s. Rather, they labeled them as English /u/s. Such a perceptual pattern may not fit the Two Category assimilation either because there are not long /i:/ and short /i/ categories in the Mandarin system. Therefore, it seems that the PAM model cannot be easily used to account for the Mandarin and Cantonese listeners’ perception of /i/-/i/ in the current study. Mandarin and Cantonese listeners who made use of the duration cues to perceptually distinguish the English /i/ and /u/ did not simply “assimilate” the /u/ sound to the L1 /i/. This phenomenon suggests that assimilation is not the only strategy the L2 learners adopt in cross-linguistic speech perception. It is important to point out that the PAM Model is not a learning model. Therefore, it does not account for the possibilities of development of L2 phonetic categories over time through learning. The possible reasons for trainees’ differences in perception of the three target vowel pairs will be discussed in later sections.
The interpretation of the current data in terms of the PAM model is, at best, speculative. The real nature of assimilation may be assessed through direct measurement by mapping the L2 segments onto the L1 system through cross-linguistic perceptual tests on monolingual Mandarin and Cantonese speakers. In other words, the perceptual test should involve monolingual Mandarin and Cantonese speakers labeling English vowels in terms of Mandarin and Cantonese vowels. Several studies using the direct measurement method to assess the assimilation patterns have been reported in recent years (Flege et al., 1997; Munro et al., 2001; Strange et al., 1998).

Apparently, both the perceiver and the phonetic differences between L1 and L2 vowel systems can affect or determine the assimilation patterns. In other words, how an L2 phone is assimilated to a certain L1 category is related to both phonetic similarities and differences between L1 and L2 categories and the way individual learners actually perceive them. Flege’s Speech Learning Model seems to place more emphasis on the role of the learners and their “perceived differences” between L1 and L2 sounds. The SLM states that L2 categories can be established when the differences between L1 and L2 sounds or between two different L2 sounds are perceived. Under this view, the subjects who perceived the /ɪ/-/i/ contrast as two different categories but the /ʊ/-/u/ and /e/-/æ/ both as single categories can simply be viewed as perceiving the differences between the vowel pairs. Of course, the phonetic differences between the three target vowel pairs affect the way listeners perceive these vowel contrasts. The important point is that phonetic differences may not be sufficient to characterize the different perceptual patterns observed with the /ɪ/-/i/ and /ʊ/-/u/ contrasts in the current study. This point can also explain the commonly observed phenomenon that
individual speakers of a same L1 may have different perceived phonetic categories for L2 phones that do not contrast in their L1 system.

Furthermore, the trainees’ significant improvement in perception of both the synthesized and the natural tokens of the three vowel contrasts as a result of training seems to support the claim of the Speech Learning Model that perceptual categories can be established if listeners can perceive the differences between the target vowel contrasts. Intensive perceptual training under laboratory conditions appears to be effective in the establishment of new phonetic categories, although Flege seems to be concerned primarily with the natural acquisition environment. In addition, phonetic categories can also be modified if the non-native speakers learn to use the correct or native-like cues for distinguishing non-native contrasts through training. As in the case of the /i/-/ɪ/ contrast in the current study, the trainees learned to shift their attention away from duration cues to spectral cues through training. Therefore, their perceived category distinction for the /i/-/ɪ/ contrast after training was different from what it was at pretest, although they did appear to have a category distinction for this particular pair based on duration in the first place. The current data in perception learning provided support for the Speech Learning Model.

5.2.2. Other Factors that Influence Perception Patterns

The differences in perceived single or two-category distinctions across the three vowel pairs may be, in part, due to pronunciation teaching and learning in English language classrooms. The /i/-/ɪ/ contrast is commonly taught as two vowel categories that differ in length in China (Wang & Munro, 1999). In the current study, some trainees reported that their English teachers in Mainland China, Hong Kong, and Taiwan taught them long /i:/ and
short /iː/ for the /iː/-/uː/ contrast. In contrast, the participants were generally not aware of the
duration differences between the /uː/-/oː/ and /æ/-/æː/ categories as the pairs were not taught to
them as long versus short vowels. In other words, pronunciation instruction in the foreign
language classroom was misleading, as in the case of /iː/-/uː/ contrast, and may have affected
the speakers’ perceptual patterns for L2 vowel contrasts. Moreover, it appears to have drawn
their attention to only one of the three contrasts. They seemed to be aware that the /iː/-/uː/
distinction was indeed an English contrast, but may not have realized that the other two
distinctions even existed.

Another possible explanation for the differences in learning between the /iː/-/uː/ and
/uː/-/oː/ distinctions may be related to the difference in functional load each of the three target
vowel contrasts bears. The /iː/-/uː/ distinction bears more functional load and has more
minimal pairs in English than does the /uː/-/oː/ contrast (Brown, 1988). Hardly any commonly
used vocabulary items are distinguished by the /uː/-/oː/ contrast in English (Brown, 1988;
Kucera & Francis, 1967). This is clearly reflected in the extremely small number of minimal
pairs found in English for this training experiment. Therefore, in real communication, the
confusion caused by the substitution of /iː/ for /uː/ would cause more problems than the
substitution of /uː/ for /oː/. The greater need to contrast /iː/-/uː/ in communication probably
results in more painstaking efforts from learners in finding the acoustic properties that signal
the difference between the vowels in order to understand and to be understood. This might
provide an explanation for the single category perception of the /uː/-/oː/ pair.
However, the training data also showed that the overwhelming majority of trainees had the most difficulties with the /ɛ/-/æ/ pair and spent more time and training sessions on this pair than the other two vowel contrasts. This phenomenon cannot be explained by the functional load differences because, compared with the /u/-/ʊ/ pair, the /ɛ/-/æ/ contrast actually has a much greater functional load and more minimal pairs (Brown, 1988). The need to distinguish the /ɛ/-/æ/ contrast did not (at least not in the current study) appear to help the learners to find strategies to distinguish the contrast. In fact, this pair appeared to be more resistant to perceptual learning.

One possible reason for the trainees’ special difficulties with the /ɛ/-/æ/ pair may be related to the high degree of spectral overlapping for the /ɛ/-/æ/ contrast. Evidence from L1 vowel perception tests suggests that the American English /ɛ/-/æ/ contrast is intrinsically more spectrally confusing than the /i/-/ɪ/ and /u/-/ʊ/ contrasts. Hillenbrand & Clark (2000) reported that native American English listeners’ identifications of English /i/-/ɪ/ and /u/-/ʊ/ contrasts were not affected by distorted duration (edited long and short vowels) differences because spectral differences were sufficient for the differentiation of these vowel pairs. In contrast, American English listeners responded to duration contrasts for the duration-altered /ɛ/-/æ/ contrast. The authors speculated that the /ɛ/-/æ/ contrast showed a greater degree of spectral overlap than the /i/-/ɪ/ and /u/-/ʊ/ contrasts. Of course, these findings for American English may not automatically be used as an explanation for the problem of the /ɛ/-/æ/ contrast in the current study because the Canadian English /ɛ/-/æ/ contrast might be phonetically different from the American English counterpart. More evidence is needed for a
firm claim of a greater level of difficulty in perception of certain vowel contrasts than others because of intrinsic vowel spectral overlapping.

5.3. RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION

The term “transfer” of perceptual learning to production (Bradlow et al., 1997) is commonly used in reporting training studies. In general, significant changes in trainees’ productions at post-test that are not seen with the control group are viewed as a transfer of perceptual learning to production if the trainees take only perceptual training.

Previous studies have yielded evidence that non-native speakers’ perceptual learning through perceptual training on L2 consonant contrasts (Rochet, 1995; Bradlow et al., 1997) and L2 lexical tones (Leather, 1997) can transfer to production. In those cases, the trainees’ performance in production improved significantly from pretest to post-test with training only in the perception mode. Although the effect of training on perception in the current study was comparable to the effects observed in previous studies, the current study did not match the success of those studies in terms of transfer of learning to the production mode. The trainees’ improvement in perception was not accompanied by significant improvement in production. This gives rise to the question of why such a difference was observed. Although no definite explanation can be offered to account for the differences in the learning of L2 consonant and vowel contrasts, one possible reason may be proposed. It is speculated that the lack of fixed places of articulation for vowels as compared with consonants might be a factor. The movement of the tongue during the process of vowel production might make it difficult to pinpoint the exact place of articulation of a vowel contrast, causing problems in making references to the perceived vowel differences when attempting to produce them. In a similar
study on Japanese speakers’ perceptual learning of American English vowels, Yamada et al. (1998) reported that after 45 sessions of perceptual training, the subjects who made significant progress in perception did not show signs of improvement in the production mode. The authors attributed the lack of improvement in production to the differences in vowel and consonant productions because articulatory control for vowel production is continuous while the articulatory control of consonants is somewhat discrete (Yamada et al., 1998).

The current findings suggest that learning in perception does not bear a straightforward relationship to learning in production. The relationship between L2 speech perception and production is a complex issue that may not simply be accounted for by any single theory. Although the current study was not designed to examine the theoretical bases underlying the relationship between perception and production, the findings have provided empirical data on this issue. Two relevant theories, the Motor Theory or the Direct Realist Approach, were introduced in Chapter 1. Both emphasize the direct link between perception and production.

The current data did not provide evidence to support the theoretical accounts of either the Motor Theory or the Direct Realist Approach in speech production and perception. According to the Motor Theory (Liberman, 1991; Liberman & Mattlingly, 1985), the intended gestures (neuromotor commands) used in the production of certain speech sounds are recovered and references to these gestures are made when a listener tries to decode these sounds in speech perception. Under this view, the changes in perception should be accompanied by simultaneous changes in the production mode as references to the acquired

---

1 It is important to note that training techniques and evaluation methods used in those studies are not the same as the ones used in the current study. For example, the production and perception data in the current study were
gestures or articulatory commands at the phonetic level would be made. In other words, the articulatory commands would be modified at the same time through perceptual learning.

The Direct Realist Approach posits that the actual gestures produced by the vocal tract are perceived directly. The gestures are, therefore, directly detected in speech as they are directly presented in speech (Best, 1995). Under this view, perceptual learning is learning the gestures of the production itself. The changes in perception through perceptual training should be accompanied by simultaneous changes in production without even going through the phonetic module mediating between perception and production. The current data provide no support for the mechanisms described in either approach. There was at least a gap between perception and production on the three vowel contrasts under investigation.

Flege’s Speech Learning Model also predicts that the learners will be able to produce an L2 sound when its phonetic category is established. According to Flege, “the production of a sound eventually corresponds to the properties represented in its perceptual phonetic category representation” (Flege, 1995, p. 239). Although the SLM makes such a prediction of a direct relationship between perception and production, it does not state clearly how the establishment of categories leads to improved production. This statement does, however, imply that perception leads production. The word “eventually” suggests that learning in perception and production may occur at different paces. The current data did not give full support for this prediction on production. At least the perceived differences between the vowels did not automatically or immediately lead to success in producing the differences for many trainees. There was no evidence of better production three months after the training was completed either (see Chapter 3 for details).

5.4. CONTRIBUTIONS AND LIMITATIONS OF THE CURRENT STUDY

analyzed with the same statistical methods while those in the Bradlow et al. (1997) study were not.
Most perceptual training studies of L2 phonetic learning have focused on non-native consonant contrasts. Very few studies have dealt with modifying adult L2 learners’ perceptions of non-native vowel contrasts. The current study has provided important empirical data for L2 vowel training under laboratory conditions.

Another contribution of the current study was the exploration of efficient training techniques in terms of using both natural and synthesized training stimuli. The results of the current study provide evidence that the combined use of synthesized and natural tokens presented in a way that increased the variability of both stimuli types can be effective in the perceptual training of non-native vowel contrasts. In addition, the mixed-talker presentation of natural words produced by multiple talkers in each training session increased the variability of talkers and might have helped to reduce the talker effect commonly reported in previous studies (Logan et al., 1991; Lively et al., 1994).

Several limitations of the current study must be pointed out. One problem with the experiment design was the different amount of time each trainee spent on training. However, this problem may be viewed differently or even as a strength from a different perspective. First, the amount of training time in a training session and duration of training in a long-term training study has no fixed standard, a factor that raises complex issues. In a short-term or one-session training experiment, performance measures rather than amount of training time is sometimes used as justification for the termination of training. For example, in Wang and Munro (2000), each trainee’s identification scores of the target contrast under training had to reach 95% before the training stopped, regardless of the amount of time it took the trainee to reach the standard. For long-term training (training that occurs over several days or weeks), very few studies have set any criteria as justification for the termination of training. Many
studies seem to have had a predetermined number of sessions of training for all trainees (Rochet, 1995; Yamada et al., 1998), and post-tests were conducted as soon as the training sessions ended.

The learner-centered and self-paced training adopted in the current study helped to keep the unpaid trainees motivated and allowed each trainee to address specific problems identified by the pretest. Therefore, this training procedure had its advantages. On the other hand, having all the trainees taking the same amount of training time or the same duration of training has the advantage of controlling the treatment factors in an experiment design. There is a trade-off between the two concerns in determining the training time: the motivational issue in keeping the trainees committed and the control of treatment elements in a training experiment. Either choice has its limitations. For example, previous training studies that controlled the number of training sessions have not reported how the trainees were motivated and maintained throughout the period of training (Bradlow et al., 1997; Yamada et al., 1998).

A related issue was that the number of training sessions each trainee had on each of the three target vowel contrasts was not equal. This was in part due to the fact that only a few trainees had serious perceptual problems with all three target vowel contrasts as determined by the pretest. As the majority of trainees had perceptual difficulties with only one or two pairs under training, most of them chose to address their own problems by focusing on one or two pairs. In spite of the fact that the effect of training on each vowel contrast could not be assessed in isolation, as training on one pair may have some impact on the other two pairs, the differences in training sessions across three target vowel contrasts could be a problem for direct comparison of the effect of training on learning across vowel pairs. Therefore, it would be ideal to include participants who had problems on all three vowel contrasts under
investigation so that each trainee could have the same number of training sessions on each vowel contrast. However, it is important to point out that this ideal condition may still not be easily established because the /e/-/æ/ contrast took more training sessions to learn than the other two pairs for many trainees.

For a well-designed laboratory training experiment, the treatment procedures should be controlled and balanced. In this respect, a drawback of the design of the current study was the unequal number of participants in the trained and the control groups. More subjects were assigned to the training group at the initial stage of training with the aim of maintaining as many trainees as possible for the trained group. It would certainly be more desirable to obtain more control subjects for estimation of the training effect. A more serious problem is the fact that ceiling effect was observed with some trainees on some target vowel pairs at the beginning of the training. Because of this, a lack of improvement on one or two of the three vowel pairs under training may not indicate that training was not effective for these participants.

Finally, the lack of evidence of a transfer of perceptual learning to production of the three target vowel pairs under training suggests that training in perception alone may not be sufficient. Future studies should explore a combination of production and perceptual training for better results.