CHAPTER I

AN OVERVIEW OF CROSS-LINGUISTIC SPEECH PERCEPTION

1.1. INTRODUCTION

It is well known that second language (L2) learners often have difficulties with perception and production of non-native phonological segments that either do not occur or are realized differently in their first language (L1). This problem can occur even with highly proficient, fluent L2 speakers with a long period of residence in the ambient language (Flege, Mackay, & Meador, 1999; Munro, Flege, & MacKay, 1996). However, there is evidence that L2 learners can improve their perception and production of non-native sounds through intensive laboratory training (Bradlow, Pisoni, Yamada, & Tohkura, 1997).

This thesis explores the effect of perceptual training under laboratory conditions on the learning of non-native vowel contrasts. Native Mandarin and Cantonese speakers who are highly proficient in English will participate in a long-term training program for learning the English /i/-/u/, /u/-/u/, and /e/-/æ/ contrasts that pose serious problems for speakers of Chinese languages (Hung, 2000; Munro, Wang, & Li, 2001; Wang, 1998; Wang & Munro, 1999).

The thesis consists of five chapters with the following layout. Chapter 1 is an overview of current theories and empirical studies of L2 speech perception and production. The discussion includes the language-general nature of infant speech perception, the language-specific nature of adult monolinguals’ cross-linguistic speech perception, and the difficulties adult L2 learners face in the perception of non-native speech sounds. The review also includes two influential L2 speech-learning theories, the Speech Learning Model and the Perceptual Assimilation Model. Although the current study is not intended to explore learning models or to test a particular theoretical framework, the literature review in the first
chapter examines the influence of the first language experience on the learning of second language speech sounds. A number of other factors including age of learning and L2 learning experience are also discussed.

Chapter 2 reviews studies of perceptual training under laboratory conditions. Success in the perceptual learning of non-native segmental contrasts through laboratory training is directly related to variables such as the background of the trainees, the types of contrasts, and the training procedures adopted. Various training techniques used in previous studies are introduced and compared for their effects on learning. The discussion covers four areas, a) experiment design, b) training procedures such as token presentation, task type, and the duration of training, c) assessment of the effect of training, and d) the limitations of previous studies that motivate the current study. A brief description of the Mandarin and Cantonese vowel systems is also provided, and comparisons of the English vowel pairs /ɪ/-/ɨ/, /ʌ/-/ʊ/, and /ɛ/-/æ/ with the Mandarin and Cantonese vowels are made. Finally, the design of the current study is presented and the specific research questions to be addressed are identified.

Chapter 3 reports the design, implementation, and results of the long-term perceptual training experiment for learning the English vowel contrasts /ɪ/-/ɨ/, /ʌ/-/ʊ/, and /ɛ/-/æ/ in the perception mode. Two training methods are adopted: the fading technique with synthesized stimuli and the high variability method with natural tokens. The pretest - training - post-test procedure is followed by a retention test three months after the training is completed to assess the long-term effect of the training.

Chapter 4 reports the results of two evaluation procedures - an intelligibility test and acoustic measurement - that assess the effect of the perceptual training described in Chapter 3 on production. In the intelligibility test, Mandarin and Cantonese subjects’ productions of the
target vowel pairs at pretest and post-test are identified by phonetically trained native English
speakers to determine whether the trainees’ intelligibility in production has improved as an
effect of training. The second test involves the measurements of vowel durations on the
speakers’ productions of the three target vowel pairs to assess whether training promotes
more native-like production of duration differences between English tense and lax vowels.

Chapter 5 summarizes the findings and discusses the implications of the current
study. The goals of the study and the research questions are revisited in a discussion of
general findings. Contributions and limitations of the current study are summarized as well.

1.2. STUDIES OF INFANT SPEECH PERCEPTION: LANGUAGE-GENERAL

PERCEIVERS AND DEVELOPMENTAL CHANGES

1.2.1. Introduction

In recent decades, empirical studies of cross-linguistic perception and of infant speech
perception have led to theories that increase our understanding of adult L2 speech learning
problems. Current theories of L2 speech learning have hypothesized that adult L2 speech
perception problems are due to assimilation of L2 segments to L1 categories. The nature of
such assimilation patterns has been characterized in more than one way (Best, 1994; Flege,
1995a). In order to have a better understanding of L2 speech learning problems, it is
necessary to examine the related research in infant speech perception, adult monolingual
speakers’ speech perception, L2 learners’ cross-linguistic speech perception and production,
and the relationship between L2 speech perception and production. Other factors that are
believed to influence L2 speech learning, such as age of learning, L2 experience, and
phonetic salience will also be discussed.
1.2.2. Infant Speech Perception

Research in infant speech perception has led to the consensus that infants can perceive speech sounds, particularly consonants, categorically (Aslin, Pisoni, Hennessy, & Perey, 1981; Lasky, Syrdal-Lasky, & Klein, 1975). Furthermore, infants are argued to be language-general perceivers at the phonetic level as speech sounds are perceptually differentiated regardless of their phonological functions in different languages. Starting in the later six months of the first year of life, this universal perceptual pattern undergoes profound change due to increasing experience with the ambient language (Best, 1994; Polka & Bohn, 1996; Strange, 1995; Werker, 1994; Werker & Polka, 1993).

Studies of infant speech perception have focused mainly on consonants. For example, Best and her colleagues found that infants in an English environment could hear the Zulu apical vs. lateral click contrast /!/ and /l/ (Best, McRoberts, & Sithole, 1989). Spanish learning infants were able to discriminate the English voiced and voiceless contrast even though it is not used phonemically in Spanish (Lasky et al., 1975). Werker and her colleagues found that English learning infants of 6-8 months old, but not 10-12 months old, were able to discriminate the Hindi voiceless unaspirated and breathy voiced /\textipa{t}\textipa{a}/-\textipa{\textipa{d}ha}/ contrast, the voiceless unaspirated dental and retroflex /\textipa{\textipa{t}\textipa{a}}/-/\textipa{\textipa{t}\textipa{a}}/ place distinction, and the Thompson glottalized velar and uvular /k’/ and /q’/ contrast (Werker, Gillbert, Humphrey, & Tees 1981; Werker & Tees, 1984). Young native English-speaking children at the ages of 4, 8, and 12 performed poorly, just as the English adults did when tested on these contrasts (Werker & Tees, 1983). Based on these studies, the authors concluded that the decline in infants’ ability to perceive non-native consonant distinctions begins late in the later six months of the first year.
Of the few studies of infant vowel perception in recent years, findings have not always been in agreement with those of consonant studies. In one study, Best (1999) tested English learning infants on the Norwegian high front vs. central unrounded vowel contrast that adult speakers of American English had no difficulty contrasting. Infants of 3-5 months discriminated the vowels well, but the 10 month olds had a complete lack of discrimination. In another study, Polka and Werker (1994) examined English infants’ perception of the German vowel contrasts /u/-/y/ and /o/-/y/. They found that the 4-6 month olds could perceptually differentiate these vowel contrasts while the 6-8 month and 10-12 month olds did not perform as well. However, in a later study (Polka & Bohn, 1996; see also Bohn, 1998), both German and English infants of 6-8 and 10-12 months were tested on English /det/-/dæt/ and German /dut/-/dyt/ contrasts. No differences were found between the two language groups on either the German or English vowel contrasts, nor were there differences across the age groups.

The findings of Polka and Bohn’s (1996) study did not fall in line with the results of the consonant studies on which the conclusions of pre-lingual infant speech perception were based. Yet, Bohn (1998, p. 9) pointed out that “the absence of any differences between the age groups and the language groups in the Polka and Bohn (1996) study suggests that some non-native speech contrasts may not be susceptible to language-specific processing bias by the end of the first year of life”. The issue is further complicated by the fact that vowels and consonants may be perceived differently in early infancy. Whether the perception of vowels by infants at this stage is categorical or not needs further exploration, as we know that vowels tend to be perceived more continuously than consonants (Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969; Polka & Bohn, 1996). More studies with different vowel
contrasts and language groups are needed before any conclusions can be made on infant perception of vowels.

In addition to segmental tests, studies of infant speech perception also explored prosodic features including tones, stress and intonation. The results suggest that in the second half of the first year in life (Jusczyk, Cutler, & Redanz, 1993), or even earlier (Mehler et al., 1988), English-learning infants began to show sensitivity to the structure and organization of native language sound patterns and demonstrated sensitivity to native prosodic properties. The findings are consistent with the results based on the consonant studies.

To summarize, the majority of studies of infant speech perception, especially those based on consonants and prosody, conclude that infants have the ability to perceive speech sounds at a language-general level in the first six months of life. Increased experience with the ambient language in the later half of the first year causes a decline in infants’ ability to perceive sounds in a language-general manner. It is the specific language experience that shapes the way they categorize the speech sounds in a phonologically relevant way.

1.2.3. The Nature of Developmental Change: Reorganization of Attention

The shift from language-general to language-specific in perceptual patterns is now viewed as attentional reorganization for speech functions rather than as a “loss” of sensory abilities in adulthood (Best, 1994; Polka & Bohn, 1996; Strange, 1995; Werker, 1994; Werker & Polka, 1993). Best argued that the effect of the native language on perception of non-native contrasts is “neither absolute nor permanent, and hence, cannot be fully accounted for by sensory-neural mechanisms” (Best, 1994, p. 173). There is evidence that not all non-native segmental contrasts are equally difficult to perceive. For example, adult native English listeners were able to discriminate the German vowel contrasts /u/-/y/ and /u/-/y/ with high
accuracy without training (Polka, 1995). Furthermore, adults could actually differentiate the physical differences at a phonetic level when the methods of presenting these differences were used appropriately (see Strange, 1995). Studies also show that with the proper design of the perceptual tests, for instance, the shortening of the inter-stimuli interval (ISI), adult listeners were able to discriminate the contrasts at the auditory level (Werker and Tees, 1984).

Another factor that supports the view of developmental change in speech perception not as “loss” but as attention reallocation comes from the improvement in adult L2 learners’ perceptual ability as a function of increased experience in the L2 environment (Bohn & Flege, 1990, 1992; Flege, Bohn, and Jang, 1997). Brief sessions of perceptual training using synthesized stimuli proved to be effective in shifting native Mandarin listeners’ attention from duration to spectral cues in English tense–lax vowel pairs (Wang & Munro, 1999). Studies using naturally-produced tokens in the laboratory (Bradlow et al., 1997; Logan Lively, & Pisoni., 1991, 1993; Logan & Pruitt, 1995; Strange, 1992; Yamada, 1993) have also demonstrated that dramatic perceptual learning can take place during relatively short but concentrated periods of training.

The reorganization that takes place early in life when the first language experience begins and continues into adulthood is believed to facilitate first language acquisition. It also explains why adult monolingual speakers are language-specific perceivers of speech sounds and why adult L2 learners often have difficulties in perceiving and producing L2 sounds that are not found in L1. The findings of infant speech perception research shed light on adult monolingual speech perception and have a profound impact on our understanding of adult L2 speech learning.
1.3. ADULT SPEECH PERCEPTION

1.3.1. Monolingual Speakers’ Language-Specific Speech Perception

Adult monolingual speakers are believed to be language-specific perceivers of speech sounds (Bohn, 1998; Polka, 1992; Strange, 1995). Lisker and Abramson’s (1964, 1970) classic study of multiple monolingual speakers’ perceptions of initial stop consonants provided strong evidence for this claim. Their cross-linguistic study of synthesized stops with carefully manipulated Voice Onset Time (VOT) steps showed that adult speakers of different languages perceived physically identical initial stops in a way that was determined by the phonologically distinct inventories of stop sounds in their native languages. For example, speakers of two different L1s with a two-way voicing contrast would classify sounds with the same VOT values differently, according to their language experience. Both would also differ from listeners whose L1 made use of a three-way distinction for their stops. Therefore, boundaries for stop categories for monolingual speakers are determined according to listeners’ L1 inventories.

Boundary differences in the categorization of speech sounds that differ in place of articulation as a function of language-specific experience have also been reported. In a study of Japanese and English listeners’ perceptions of English approximants using a synthesized /w/-/j/ continuum, native Japanese listeners identified more /w/ items than native English listeners did as the former shifted the boundary towards /j/ compared with native American English listeners (Best & Strange, 1992). This difference in the perceptual boundary was found to be consistent with the detailed phonetic differences between the English and Japanese /w/. Specifically, the Japanese /w/ is produced without lip-rounding and is phonetically closer to /j/.
Studies of vowels have also indicated that adult monolingual speakers’ perception and labeling of L2 vowel categories are influenced by the way these vowels contrast in the particular languages they speak. For example, both English and Portuguese systems have two categories /i/ and /u/ occupying the high vowel space, while French has a three-way contrast /i/-/y/-/u/ in this area. In a study using a synthesized /i/-/u/ vowel continuum, Rochet (1995) found that monolingual Portuguese and English listeners showed different category boundaries while labeling /i/-/u/ in an identification task. This is probably because the Portuguese and English /i/ and /u/ categories differ phonetically and cover different spaces along the French /i/-/y/-/u/ continuum. The results provided an explanation for the observation that native Portuguese listeners assimilate the French /y/ to their /i/ category while native English listeners assimilate /y/ to their /u/ category.

To summarize, previous research has provided strong evidence that adult monolinguals are language-specific perceivers. Cross-linguistic perception studies show that category boundaries of synthesized sounds are often determined by the phonological inventories of a speaker’s native language.

1.3.2. Adult L2 Learners’ Speech Perception: The Influence of L1 Experience

Reduced sensitivity to language-general speech sounds in infancy by 10-12 months and the language-specific nature of adult monolinguals’ speech perception underlie the difficulties faced by adult speakers in learning the sound system of a second language. Studies of cross-linguistic perception of non-native segmental contrasts often indicate that adult listeners perceive such contrasts with difficulty. Among such studies, the most numerous and notable are those of native Japanese speakers’ problems with the English /u/-/l/ distinction (Bradlow et al., 1997; Lively, Pisoni, Yamada, R., Tohkura, & Yamada, T. 1994;
Logan et al., 1991; MacKain, Best, & Strange, 1981; Miyawaki et al., 1975; Sheldon & Strange, 1982; Yamada, 1993; Yamada, 1999). Problems in perception of non-native place of articulation contrasts have also been reported on native English listeners’ perception of the Hindi dental versus retroflex stops: /t̪a/ and /t̪a/; Thompson velar and uvular ejectives /k’/-/q’/, (Polka, 1991, 1992; Werker et al., 1981; Werker & Tees, 1983, 1984), and Japanese listeners’ perception of English /w/-/j/, /w/-/u/ contrasts (Best & Strange, 1992).

Problems with perception of non-native voicing contrasts are also common (Rogers, 1997; Flege, 1989; Morosan & Jamieson, 1989; Werker et al., 1981; Rochet, 1995). Studies of L2 vowel perception have shown that adult speakers have problems in perceiving non-native vowel contrasts (Baker & Trofimovich, 2000; Bohn & Flege, 1990; Famoso, Schwartz, & DaSilva, 1998; Flege & Bohn, 1989; Flege, 1997, Flege et al., 1997; Flege et al., 1999; Gottfried, 1984; Gottfried & Beddor, 1988; Ingram & Park, 1997; Munro, 1993; Munro et al., 1996, Polka, 1995; Strange et al., 1998; Wang, 1997, 1998; Wang & Munro, 1998, 1999).

Adults’ problems in perceptual learning of non-native perceptions have been characterized in more than one way and are believed to be related to a number of factors. One important factor that is known to influence cross-linguistic perception is L2 learners’ linguistic experience with their first language. Cross-linguistic speech perception studies have shown that adult L2 speakers often assimilate L2 sounds to relevant L1 phones in L2 speech perception and production.

For example, in a series of studies, Wang (1997, 1998) found that native Mandarin speakers residing in Canada had serious problems in perceiving and producing English
vowels. In both perception and production tasks, they often substituted the L1 vowel categories for English vowels that do not have Mandarin counterparts.

In a more recent cross-language vowel study of native Korean speakers’ perception and production of American-English vowels, Baker and Trofimovich (2000) found that Korean late arrivals to the United States with a relatively long residence of 10 years in the L1 environment had problems both perceiving and producing the English vowels /i/-/u/ that do not have Korean counterparts and that the substitution of L1 vowel categories for L2 vowels was also common.

In another study of Korean speakers’ perception of Australian English vowels, Ingram and Park (1997) found that a phonemic merger of the Korean vowel pair /e-æ/ in the younger generation influenced their perception of the similar /e-æ/ contrast in their L2 (Australian English). The two-to-two assimilation of /e-æ/ for the older Korean speakers vs. the two-to-one assimilation for the younger generation had a different impact on the listeners’ perception of the similar vowel pair in the L2.

Beddor and Strange (1982) investigated the effect of linguistic experience on perceptual categorization of the oral-nasal distinction in consonants and vowels by two L1 groups. Native Hindi and American English speakers were tested on identification and discrimination of synthesized [ba-ma] and [ba-bā] series. As expected, no language-group differences were found in the discrimination of the consonant series [ba-ma] (a phonemic contrast in both languages) as nasal consonants were perceived categorically in both languages. However, on the vowel series, [ba-bā] (phonemic only for the Hindi speakers), cross-linguistic differences in discrimination were found. Hindi speakers’ perception of the oral-nasal distinction was categorical while native English speakers’ perception of the [ba-
bā] vowel series was more continuous. The latter group accurately discriminated the differences not only across categories, but also within the oral category.

The possibilities of positive transfer of L1 experience have also been examined. Some studies have explored whether the presence of certain features of place of articulation in the first language would assist its speakers in discriminating the non-native consonantal distinctions that share the same place feature but differ in manner. For example, Polka (1991) examined whether Farsi speakers would have an advantage over English speakers in learning the Salish glottalized velar and uvular stop contrasts which form a non-native distinction for both native English and Farsi listeners. For the Farsi listeners, there is a phonemic feature distinction, velar versus uvular, occurring in different manners, voiced stops and voiceless fricatives. English, however, does not exploit the uvular place in its consonant system. The hypothesis was that if the Farsi made use of the uvular place feature to identify the Salish distinction in a different manner of articulation, they would perform better than the native English listeners. However, the Farsi listeners as a group did not demonstrate a perceptual advantage because of their experience with the place contrast in another context (Polka, 1991). The results suggested that the uvular place distinction in the L1 system was embedded in a certain manner of articulation for the particular consonant contrasts and did not transfer to a segment with a different manner of articulation.

Similarly, voice features that are believed to be less resistant to learning than place features in cross-language speech perception, and relatively easier to acquire through training (Strange, 1992) do not appear to be transferable cross-linguistically. For example, native Canadian French speakers found it difficult to distinguish the English /ʊ/ and /θ/ voicing contrast (Jamieson & Morosan, 1986) although in French, voice distinctions are found in its
fricative pairs like /s/-/z/, /f/-/v/. Voicing features that are shared by L1 and L2 on certain segments may not be transferred to new L2 sounds automatically. Consequently, the sharing of the same features, either place or voicing of L1 and L2 segments, may not necessarily facilitate learning.

There is also evidence that speakers’ first language experience may determine the way certain phonetic cues are used in cross-linguistic speech perception (Bohn, 1998). For example, native English speakers use primarily spectral cues to differentiate vowel contrasts (Munro, 1993; Strange et al., 1998) even though duration differences co-vary with spectral differences in English tense-lax vowel contrasts (Gottfried & Beddor, 1988). On the other hand, speakers whose L1 exploits duration differences to contrast vowels will attend to vowel duration in cross-linguistic vowel perception. For example, the difference between the high, front, unrounded tense-lax vowel pair was heard as /i/-/i:/ distinction by native Arab speakers but as /i/-/u/ by English speakers (Munro, 1993).

Evidence from empirical studies has led to theories or models of perceptual assimilation. These theories hypothesize that difficulties perceiving L2 distinctions are due to L2 segments being assimilated to L1 categories. The most influential models in recent decades are Flege’s Speech Learning Model (SLM) and Best and her colleagues’ Perceptual Assimilation Model (PAM). Both models attempt to account for how the L2 sounds are related to and assimilated to the L1 system at different levels (phonemic, phonetic) to predict levels of difficulty in learning the nonnative contrasts.

**1.4. SPEECH PERCEPTION MODELS**

**1.4.1. The Speech Learning Model**

production depends on the establishment of new phonetic categories (long-term memory representations) for the L2 segments. The early version of the SLM states L2 phones that are “new” (greater dissimilarity between L1 and closest L2 sounds) will eventually be mastered if learners can perceive the phonetic differences between the L1 and L2 phones. On the other hand, L2 phones that are “similar” to L1 categories will pose more difficulties because “equivalence classification” will block adult L2 learners from establishing a phonetic category for the “similar” but not for the “new” L2 phones. Moreover, the SLM is a dynamic model in the sense that it also predicts the change in learners’ perceptual patterns as a function of increased L2 experience in the course of learning.

There is some empirical support for the SLM. For example, Flege (1987) found that native English speakers produced the “new” French phone /y/ successfully but not the “similar” phones /u/ and /t/. There is also evidence that experienced learners of an L2 demonstrated more native-like perceptual patterns of L2 vowels than less experienced learners (Bohn and Flege, 1990, 1992).

However, a number of problems have arisen for the Speech Learning Model. The earlier version of the SLM has been criticized for lacking explicit criteria for the classification of “similar” and “new” phones, especially vowels (Blankenship, 1991; Ingram & Park, 1997). It is not clear as to how “close” the L1 and L2 vowels should be in order to be classified as “similar” or “identical”. Another problem with the SLM is that it cannot account for the complicated results of many empirical studies in the area of speech learning. For example, a study of native Italian speakers’ productions of English vowels, (Munro et al., 1996) found the hypothesis that English vowels which are acoustically “close” to Italian vowels should be less well produced than vowels that are more distant was not confirmed.
In more recent years, instead of the old categorization of “new” and “similar” phones cross-linguistically, Flege has adopted the term of “perceived phonetic differences” between the L1 and L2 phones or two L2 phones. The modification seems to place more emphasis on the role of the perceiver (depending on how much the L2 learner hears the phonetic differences between the target contrasts) than on the nature of the acoustic differences between the two sounds. The emphasis on the listeners or perceivers in terms of “perceived phonetic” distance would also provide better explanations for learner differences in the same L1 group learning the same L2 contrasts. For example, certain phonetic differences between an L1 and L2 phone may be perceived by some but not by other learners from the same L1 speaking group.

Another important point of the SLM is that it emphasizes the relationship between perception and production by assuming that accurate perception of the phonetic differences between two L2 phones will eventually lead to the correct production of these differences. Although this prediction still needs to be verified by empirical data, such a prediction of a link between perception and production certainly distinguishes this model from other pure perception models, such as Best’s Perceptual Assimilation Model.

1.4.2. The Perceptual Assimilation Model

Best’s Perceptual Assimilation Model (PAM) (Best, 1994; Best, McRoberts, & Sithole, 1988; Best & Strange, 1992) also predicts the levels of difficulty in differentiating the L2 segments on the basis of how the L2 sounds are assimilated to L1 segments. The PAM model goes beyond the range of one-to-one segment comparisons between the L1 and L2 systems and states that two members of a non-native contrast can be assimilated to native phones in a number of different ways. For instance, the two L2 sounds could be assimilated
to two different L1 phonemes that are similar to the nonnative contrast, the Two Categories type (TC). The non-native pair could also be assimilated to a single native phoneme equally well or poorly, the Single Category type (SC), or to a single non-native phoneme with different phonetic match which will result in one better assimilated than the other, the Category Goodness type (CG). In addition, the PAM model attempts to account for the situation of non-native speech sounds that are unassimilable to any L1 phonemes. At this level of perception, according to Best, assimilation does not occur, as there is no organization of L1 and L2 sounds at phoneme level.

The PAM also predicts that TC type will be highly accurate and least difficult, the SC the most difficult and the CG in the middle range, from quite difficult to easy depending on the degree of “goodness” between the two phones perceived. The unassimilable L2 sounds should also be easy to differentiate, depending on the perceived auditory or phonetic difference independent of the L1 phonemic system.

Support for the PAM model has been found in a number of studies. For example, in Best & Strange (1992), the American English /w/ and /j/ appeared to be assimilated to the corresponding Japanese /w/ and /j/, a typical case of Two Category assimilation. The English /\l/ and /l/ sounds were often believed to be assimilated to a single Japanese /r/, a Single Category type of assimilation (Best et al., 1988). Polka and Werker (1994) used naturally-produced German front and back rounded vowel pairs to test both native English adult speakers and English born infants’ perceptual patterns. The German front and back rounded vowels both appeared to be assimilated to the English back rounded vowels but the ratings showed the front vowels were assimilated less well to the English vowels. Thus, the authors claimed, the English listeners perceived clear differences in category goodness for the
German vowels within each contrast. The high level of accurate performance on non-native vowels as shown in this study was taken as support of the Category Goodness assimilation pattern of the PAM model.

In summary, both the SLM and PAM models hypothesize that adult L2 speech perception problems are due to the assimilation of L2 segments to L1 categories although the nature of such assimilation is characterized differently. Although the current study is not designed to test the assimilation models, the results of the tests and training will be discussed in relation to both models.

1.5. OTHER FACTORS THAT INFLUENCE NON-NATIVE SPEECH PERCEPTION

It is important to point out that L2 perception is a complex process that involves many other learner factors in addition to the influence of a learner’s first language experience. One cannot fully address the learner factor in L2 speech perception and production without taking into consideration the age of learning. It is commonly believed that children are more successful than adults in learning a second language in the aspect of phonology (Singleton, 1995). Some researchers have proposed the Critical Age Hypothesis (CAH), which states “foreign accent cannot be overcome easily after puberty” (Lenneberg, 1967, p. 176). Over the past two decades, the debate on the CAH has not led to complete agreement. During this period, empirical studies examining the effect of age of arrival in the L2 environment and age of learning on the level of success in pronunciation have not reached consensus on the exact age of learning a second language at which a speaker’s accent will be detected (Asher & Garcia, 1969; Flege, 1988; Flege et al., 1999; Patkowski, 1990; Thompson 1991; Tahta, Wood & Loewenthal, 1981). Although researchers do not agree on whether there is a “critical age” in foreign language speech learning, they have come to the agreement
that “younger equals better” in L2 speech learning (Singleton, 1995). Therefore, the degree of success in learning non-native segmental contrasts is related to the learner’s age of learning.

L2 experience is believed to be another important factor that influences L2 speech learning. Adult learners’ problems with non-native speech contrasts do not always remain unchanged in the course of learning. Empirical studies indicate that learners with increased experience in the target language in the L2 environment outperform less experienced learners from the same L1 background. (Bohn and Flege 1990, 1992; Best and Strange, 1992)

Phonetic salience, or the psychoacoustic factor, is also believed to influence L2 speech perception. Burnham and his colleagues (Burnham, Earnshaw, & Quinn, 1987) have proposed that the degree of difficulty at the segmental level can be predicted on a psychoacoustic basis. They categorized contrasts that are perceived with a clear psychoacoustic basis as “robust” and those perceived with a less clear psychoacoustic basis as “fragile”. According to Burnham et al. (1987), the “robust” contrasts occur in a wide range of the world’s languages while “fragile” sounds are not common among languages. Although the robust vs. fragile proposal is based on evidence from their findings in infant speech perception, it has not provided clear standards for determining which contrasts are “robust” and which are “fragile” in general. In addition, the “psychoacoustically salient” contrasts (Best et al., 1988) may not be “universal” as the case with Zulu clicks that are relatively easy to perceive proves. In recent years, the factor of phonetic salience seems to have received less attention and has not been followed by substantial studies over the past decade.

Other factors, such as a speaker’s motivation and aptitude may also play a part in the overall picture of L2 speech learning (Skehan, 1991). Although the current study is not
designed to explore these factors, awareness of them is important in our understanding of the L2 speech learning.

**1.6. THE RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION**

As the ultimate goal of learning the sound system of a target language is the success in both perception and production, perceptual learning cannot be completely evaluated without examining its relation with production. Therefore, the relationship between perception and production is an important issue in L2 phonetic learning. Although the exact link between the two parameters is not straightforward, there is no doubt some link between them.

Both the Motor Theory (Liberman, 1991; Liberman & Mattingly, 1985) and Direct Realist Model (Best, 1995) agree that there is a direct link between speech perception and production, although the two models differ in perceptual primitives and mechanisms (Best, 1995). According to the Motor Theory, the link between speech perception and production is innate (Liberman & Mattingly, 1985) and there is a common locus for speech perception and production (Liberman, 1991). The objects of speech perception are the intended phonetic gestures of the speaker. The intended gestures (motor commands) used in production of certain speech sounds (phonetic structure) are recovered, and references to these gestures are made when a listener tries to decode these sounds in speech perception. Liberman believes that there is a language module that has various components, one of which is the phonetic system (Liberman, 1991).

The Direct Realist Theory, on the other hand, 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). According to Best, the key
difference between the two theories is that the Motor Theory proposes that there is a phonetic module that represents the articulatory gestures that mediate perception and production while the Direct Realist Theory does not propose such a phonetic module. For a more detailed description and comparison of similarities and differences between the two models, see Best (1995).

Current L2 speech learning models are perceptually based. As discussed earlier, both the PAM and SLM place much emphasis on the perceptual assimilation of the L2 sounds to L1 phones. Both models relate L2 learners’ problems in L2 speech learning to the perceptual assimilation of non-native sounds to their L1 phones. According to these models, a speaker’s problems lie mostly in perception. While the PAM model does not directly predict difficulties in production, the SLM hypothesizes that perceptual learning (the establishment of a phonetic category) will eventually lead its way to production.

Empirical studies of the relationship between perception and production in L2 speech learning are still very limited and the results of such studies have not always been consistent. On the one hand, studies of both perception and production of non-native contrasts have often indicated that non-native vowels that are perceptually difficult to learn also pose production problems (Wang, 1997; Wang and Munro 1999; Frieda, Walley, Flege, & Sloane, 2000). On the other hand, Sheldon & Strange (1982) found that some Japanese learners were somewhat more successful in producing than perceiving the English /ɪ/-/ɪ/ contrast. They argued that perceptual mastery of an L2 contrast does not always precede the learners’ ability to produce “acceptable tokens” of contrasting phonemes.

The results of training studies carried out in laboratories over the past two decades provide evidence for a link between perception and production. For example, perceptual
training of native Japanese speakers on the English /u/ and /l/ contrast was found effective not only in perceptual learning, but that the perceptual learning through training also transferred to better performance in production of the contrast (Bradlow et al., 1997). Evidence from training studies on phonologically impaired children also supports the claim that perceptual training has a positive impact on production (Rvachew, 1994).

If perceptual learning through training can be transferred to production, can learning in production through production training be transferred to perceptual learning? In other words, will learning in one parameter automatically transfer to the other? In a parallel perceptual and productive training study of non-native speakers’ perception and production of the four Chinese tones at lexical level, Leather (1997) found that training in one domain only resulted in success in the performance in the other. The findings provided evidence that speech perception and production are closely related.

To summarize, both Motor Theory and Direct-Realist Approach agree there is a direct link between speech perception and production. Cross-language studies and training experiments under laboratory conditions have provided empirical evidence that the two domains are related in L2 speech learning. Although the current study is not designed to test Motor Theory and Direct-Realist models on speech perception and production, the effect of perceptual training on production of three target vowel contrasts will be assessed. The results of the current study will provide empirical data on the relation between L2 vowel perception and production.